THE EVALUATION OF SECOND LANGUAGE FLUENCY AND FOREIGN ACCENT

BY

CHEN-HUEI WU

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Linguistics
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2011

Urbana, Illinois

Doctoral Committee:

Associate Professor Chilin Shih, Chair
Associate Professor Mark A. Hasegawa-Johnson
Assistant Professor Torrey Loucks
Assistant Professor Annie Tremblay

Abstract

What is second language fluency? What is a foreign accent? Is it possible for an adult second language learner to speak fluently with a heavy accent or vice versa? What factors contribute to the perception of a fluency and foreign accent? What acoustic attributes correlate with the perception of fluency and a foreign accent?

To answer these questions, speech samples were randomly selected from two spontaneous speech corpora produced by Mandarin native speakers and learners (both heritage and English learners of Chinese). Around 400 speech samples were assessed by 43 untrained Mandarin native speakers in Taiwan. Eight rating questions that reflected the oral proficiency of fluency, nativeness, accentedness, disfluencies, pronunciation, grammar, vocabulary, and comprehensibility were evaluated. An acoustic analysis followed that examined the acoustic attributes contributing to the perception of fluency and foreign accent.

The findings of perceptual ratings showed that the disfluency rating highly correlated with the vocabulary score and vocabulary size in L2 production. This suggests that L2 fluency relates to the lack of vocabulary. The perception of accent correlated well with the pronunciation rating. Due to the strong correlations among rating variables, a principal component analysis (PCA) was submitted to reduce dimensionality and in order to reveal the internal structure of the data. The result revealed that all the rating variables contributed similar weights to the first principal component. The second principal component classified the rating variables into two categories. One group consisted of the knowledge factors of fluency, disfluency, grammar, vocabulary, and comprehensibility. The other group included the sound-related factors of nativeness, accentedness, and pronunciation. The acoustic measures loaded on the PCA rating platform demonstrated that the rate of speech and the second formant (F2) of vowels were the most powerful predictors of the perception of fluency and foreign accent, respectively.

In the vowel studies, all phonetic vowels in Mandarin in the complete syllable inventory were investigated in terms of articulation and acoustics. The findings demonstrated that Mandarin [i, ɨ, ɯ] pose different tongue positions, while the formant values of [ɨ, ɯ] are not significantly different. In addition, the consonantal contexts influenced the articulation of vowels greatly. The comparison of vowel similarities between Mandarin and English indicated that the primary difference lies in F2. Vowels in Mandarin are further back (lower F2) than that in English.

In the spontaneous data, the production of [u] by learners demonstrated the effect of L1 transfer in L2 speech. The [y] production is closer to [u], suggesting that the constraint of [+back] and [+rounded] is strong in English and difficult to disassociate. It is difficult to learn Mandarin [ɨ, ɯ] because learners need to sustain their tongue position from the preceding consonants to vowels and must learn not to move their tongues during the articulation of vowels. Non-target-like production of Mandarin [a] and [ɑ] resulted from the influences of coarticulation effects and transcription confusion. The findings suggested that segmental similarity is not the only factor in predicting L2 sound production.

This thesis integrated studies combining perceptual ratings, acoustics and articulation to demonstrate a detailed mapping relationship between speech perception and speech production. The findings advance our understanding of second language fluency and foreign accent and have implications for both language teaching and language testing.

Key words: second language acquisition, fluency, foreign accent, perceptual evaluation, spontaneous speech, vowel similarity, Mandarin vowels.

Acknowlegements

I would like to express my deep and sincere gratitude to my advisor, Prof. Chilin Shih for her tremendous support and guidance during the path of my doctoral study. Without her patience and encouragement, I would not have been able to reach this stage. She has greatly influenced my attitude toward problem solving. If not for her scientific view, sharp feedback, and valuable ideas, this thesis would not be born out.

I also thank Prof. Mark Hasegawa-Johnson for his bright and insightful comments during the biweekly project meetings over the past five years. The interdisciplinary advice of Prof. Torry Loucks has also been instrumental. I also want to thank Prof. Annie Tremblay for her constructive criticism on the aspect of second language acquisition.

I would like to acknowledge my parents for giving me the freedom to do what I want to do. Fuli, my husband accompanied me at the final stage and has been supportive of my work. My siblings have also wished me the best and have taken care of my parents while I have been away from home for seven years.

I thank my friends for their help in life and for making my stay in Champaign-Urbana full of fun and happiness — Dora Lu, Charlene Lee, Shawn Chang, Iris Lin, Carol Ehrhardt, Erica Britt, Lifeng Gu, Li-Jen Guo, Yu-Yoon Yoon, Eun-

Kyung Lee, Shen-Fu Tsai, Ai-Ting Huang, Tammy Hsu, Suma Bhat, Theresa Li, Tim Mahrt, Tae-Jin Yoon, Li-Hsin Ning, Di Wu.

This work was made possible with the Critical Research Initiatives Grant, UIUC, the NSF-funded project “DHB: An Interdisciplinary Study of the Dynamics of Second Language Fluency” under the PIs of Mark Hasegawa-Johnson, Chilin Shih, J. Kathryn. Bock, Fred Davidson, Richard Sproat and other team members. The rating survey could not be done without the 2010 Jiede Empirical Research Grant. The Study Abroad Scholarship from the Ministry of Education in Taiwan allowed me to focus on my research.

At UIUC, I thank the Image Technology Group at the Beckman Institute for their help with video processing and annotation, ATLAS for classroom data recording and for building and maintaining the fluency rating website. At U Penn, I thank Dr. Jia-Hong Yuan for providing the forced alignment program. In Taiwan, I thank Fu-Li Hsiao and Hsueh-Chun Chen for recruiting raters at National Changhua University of Education and National Chiao-Tung University. Last but not least, I thank the raters, subjects in the EMA study, and students and Chinese TAs in Chinese language classes for participating in this project.

To my parents and Fu-Li

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Background and Motivation

What is second language fluency? What is a foreign accent? Is it possible for an adolescent or an adult learner of a second language (L2) to speak with an accent, but in a fluent manner or vice versa. What factors contribute to the perception of fluency and a foreign accent? What acoustic attributes correlate with the perception of fluency and a foreign accent?

L2 learners have varying degrees of fluency and accent. Light accents hardly impact communication, whereas heavy accents may lead to inefficient communication or miscommunication. Inaccurate pronunciation, for instance, producing ‘present’ [pr`ɛznt] as ‘prison’ [prizn] causes perceptual disruptions in communication. This is primarily due to the fact that accent makes a person harder to understand. Accented speech requires more processing time for listeners to comprehend.

The presence of a foreign accent may impact a listener’s impression of a speaker’s personality, intelligence, and credibility, among others. Because of the increased difficulty in comprehension, foreign-accented speech sounds less persuasive and reduces the credibility of non-native speakers. On the other hand, listener's attitudes may also influence their interactions and their level of cooperation with non-native speakers. Studies demonstrated that listener's belief of the speaker's language background affects their perception of the speaker's pronunciation (Lindemann, 2003, 2005; Hu & Lindemann, 2009).

To avoid or reduce social conflicts resulting from foreign accent, educational strategies are needed to help learners reduce their accent and to train native listeners to improve their comprehensibility of non-native speech. Phoneme acquisition of an L2 for adults seems to be very difficult, especially for an L2 that is very distinct from one's native language (L1). Numerous studies have explored this issue and suggested that influences from the L1 transfer of phonemes, age-related factors, and daily exposure to native and target languages can have long-term impacts on accents (Birdsong, 2005; Flege, MacKay, & Piske, 2002; MacKay & Flege, 2004). Thus, a deep understanding of non-native speech is required to solve the issues from both the perspective of non-native speakers and native listeners.

This study investigates second language fluency and foreign accent in spontaneous speech produced by heritage speakers and English learners of Mandarin, as well as by native Mandarin speakers. The speech data were selected from two large corpora, namely, the Spontaneous Chinese Learner Speech Corpus and the Picture Telling Corpus, which cover varying degrees of fluency and foreign accent. Spontaneous speech may be closer to the language people speak on daily basis and thus may be more appropriate for investigating the phenomenon of fluency and foreign accent. Temporal features related to fluency and vowel production in Mandarin will be examined by measuring acoustic attributes. Listener perceptual judgements will reveal how these acoustic attributes in L2 production contribute to human perception of fluency and foreign accent. In addition, the Mandarin sound system will be introduced. Other factors relevant to L2 phonological acquisition, such as phonological distribution, coarticulation effect, articulation, and transliteration confusion are discussed. Vowel similarities between Mandarin and English are studied in order to observe the effect of L1 transfer on L2 production. The goal of this study is to advance our knowledge of second language fluency and foreign accent.

In the following section, the following research questions are addressed.

1.2 Research Questions

People perceive fluency and accent in daily life and they form impressions and operate on these impressions. The purpose of this thesis is to clarify the aspects of fluency and accent that can be tested experimentally and implemented technologically. Fluency and foreign accent are two related concepts, but they are not identical. The nature of the relationship between the two will be explored in this study. This thesis will focus on answering the following research questions:

1. What leads to the perception of second language fluency and foreign accent?

This question is addressed by collecting perceptual ratings from naïve native listeners in Taiwan on spontaneous speech samples produced by heritage speakers and English learners of Mandarin and native speakers. The questions to be rated concern aspects of *fluency*, nativeness, *accentedness*, *disfluency*, *pronunciation*, *grammar*, *vocabulary*, and *comprehensibility*. The ratings are submitted to a correlation analysis and a principal component analysis (PCA) to reveal the relationship among the variables.

> 2. What is the relationship between acoustic measures and human-rated judgements of fluency and foreign accent?

To answer this question, a number of acoustic attributes are measured, such as the number and duration of pauses, the articulation rate, the rate of speech, the phonation the time ratio, the standard deviation of vowel duration, and the average of the first (F1) and second formant (F2) values of vowels. These acoustic measures are then correlated with perceptual ratings.

> 3. How to measure vowel similarity? How similar are Mandarin and English vowels?

Most of the models of second language acquisition (SLA) make predictions based on segmental similarities between L1 and L2 sound systems. In order to test the SLA model and examine the effect of L1 transfer on L2 production, it is necessary to compare the vowel similarities between Mandarin (the L2) and English (the L1 of the learners). The data for this vowel study are collected through Electromagnetic Articulography AG500 (EMA). The articulatory properties of Mandarin vowels are investigated to understand the vowel categories in Mandarin. Then, the acoustic properties of vowels between Mandarin and English are assessed. As a result, the vocalic similarities between these two languages are established for later analysis of the corpus data.

The speech data used in previous research questions allow us to examine the aspects of vowel pronunciation which contribute to foreign accent. This examination prompts the following research questions.

> 4. How do L2 learners produce vowels in Mandarin? Do Mandarin vowels pose different levels of difficulty to L2 learners? What are the factors that influence L2 vowel production?

By investigating vowel production in spontaneous speech samples produced by native speakers and language learners, the areas of difficulty that English speakers have when they learn Mandarin vowels is revealed. Some vowels are easy to learn while others are difficult. What causes difficulty in learning them? This comparison of vowels provides an opportunity to examine the predictions of the model in SLA. Other factors influencing vowel pronunciation are also discussed.

This thesis will present data and an analysis to answer these questions and to identify the contribution of the various quantitative variables to the perception of second language fluency and foreign accents. This thesis also provide a rich data set to test the theoretical models of SLA.

1.3 Outline of the Dissertation

This thesis is organized as follows:

Following the introductory remarks, Chapter 2 first reviews previous research and discusses definitions of fluency and foreign accent. Next, the relationship between fluency and speech planning is discussed, following by contemporary theories in SLA. Subsequently, various acoustic measures and perceptual experiments relevant to fluency and foreign accent are summarized.

In Chapter 3, vowel categories in Mandarin are reviewed from the perspective of phonology, phonetics and articulation. Other factors affecting foreign accent, such as transliteration and orthographic confusion, are discussed. Previous studies that assessed segmental similarity are reviewed. In addition, the articulatory and acoustic properties of Mandarin vowels are investigated for further analysis. To show the affect of vowel transfer from L1 to L2 and to test the hypotheses in

SLA theories, vowels in Mandarin and English are compared by examining their acoustics.

In Chapter 4, the Spontaneous Chinese Learner Speech Corpus and the Picture Telling corpus are introduced. The sampling design, the experimental design of perceptual ratings and the procedure are addressed. The analysis of the corpus data, including the rating analysis, the principal component analysis, the acoustic analysis and the vowel analysis are performed and integrated to reveal the relationship between acoustic attributes and human perception of fluency and foreign accent.

Chapter 5 discusses the surface characteristics and formation of L1 fluency and L2 fluency. Issues related to evaluating segmental similarities in SLA theories and the contribution of pronunciation to foreign accent are addressed.

Chapter 6 summarizes and interprets the findings with respect to the main subjects in fluency and foreign accent in this thesis. Future studies are also proposed.

Chapter 2

Literature Review

In this chapter, I am going to survey the literature related to the topics of fluency and foreign accent. I will first present the operational definition of fluency in language testing. A speech-planning model addressing the cause of fluency and previous work on acoustic measurements of fluency will be reviewed. Secondly, I will present definitions of foreign accent used in previous work, followed by second language models that explain the interaction between L1 and L2 sound systems and how foreign accents are accounted for in these models. Finally, perception studies on foreign accent will be reviewed.

2.1 Introduction

For most second language learners, the primary goal of learning a foreign language is to communicate with people who speak a different language. From the aspect of oral proficiency, the goal of the learner is to be able to speak fluently with a lesser degree of a foreign accent while communicating with speakers in the L2 speech community. Heavy accents and disfluent speech are usually less intelligible (Wilson & Spaulding, 2010), which may result in unemployment (Hosoda & Stone-Romero, 2010), less credibility and persuasiveness (Lev-Ari & Keysar, 2010), and stereotypes and bias (Lindemann, 2005). Therefore, it is important to understand the phenomenon in non-native speech in order to design a corrective method to help language learners attain higher level of fluency and reduce their accent.

Speech produced by non-native speakers often lacks fluency and usually exhibits a foreign accent. Previous research attributes this to a variety of factors, including neurological, physiological, social and psychological factors. The neurological account is related to the onset age of learning, which determines whether a target language will be acquired as native or non-native. The best-known theory that explain the success of learning languages is the Critical Period Hypothesis (e.g. Lenneberg, 1967), which claims that there is a critical time window by which a language must be learned in order to be fully mastered. Typically, the development curve of pronunciation and grammatical system will reach a plateau where improvement becomes difficult after a certain age, while vocabulary size may continue to increase (Long, 1990). Other studies (ref. Piske, MacKay, & Flege, 2001) show that multiple factors contribute to the degree of foreign accent, including L1 background, the amount of exposure to the L2, the length of residence in the L2 environment, daily use of the native and target languages, attitude, motivation, training/instruction in the L2, the relative use of the L1 and the L2, and gender.

Many researches have studied segmental differences between L1 and L2 speech in acoustic phonetics and perception. Previous work on SLA has found that the difficulties L2 learners encounter are caused by the phonological properties of L1 and the tendency for learners to transfer their L1 system to L2 (Suter, 1976; Hammerly, 1982; Wode, 1983).

However, L1 transfer does not explain every aspect of L2 learners' foreign accents (Munro & Derwing, 1998) and L2 fluency is not straightforwardly related to L1 fluency (Derwing, Munro, Thomson, & Rossiter, 2009). Issues on fluency and foreign accent from the aspect of speech planning, acoustic-phonetic analyses, as well as perceptual evaluation, including global ratings of trained and untrained listeners and automatic testing systems based on speech recognition technology (Cucchiarini, Strik, & Boves, 2000; Yoon, 2009; Tepperman, 2009; Bhat, 2010) will be presented in the following section. Several leading theories suggest that foreign accents are related to the relationship between perception, involving auditory processing and production, involving motor control. The developed models of SLA - Native Language Magnet Model (Kuhl & Iverson, 1995), Perceptual Assimilation Model (Best, 1995) and Speech Learning Model (Flege, 1995b) will be reviewed in later sections.

2.2 Fluency

2.2.1 Definition of Fluency

This section will review the definition, cause and quantitative measurements of fluency that have been investigated from language assessment, psychology and acoustics. The term fluent literally means ‘flowing’. Hence, fluency gives the image of words flowing out effortlessly. There is no one definition of “fluency”. The notion of fluency consists of multiple dimensions, including good oral command of vocabulary, grammar, phonology and phonetics, and the ability to talk at length with few disfluencies, such as filled pauses, silences, repairs, repetitions, etc. If a speaker is an adolescent or an adult learner of an L2, he or she may be fluent but unable to acquire a native-like oral performance. A foreign language learner could speak fluently but with grammatically inaccuracies, fluently with limited vocabulary, or fluently with a heavy accent or with bad pronunciation, alternatively, he/she could speak correctly but not fluently or with a light accent but not very fluently. These impressionistic descriptions imply that fluency may be judged independently from grammar, vocabulary, accent or pronunciation. In practice, these factors are often used together to rate fluency.

In language testing, fluency is frequently used to determine a testee's level of oral proficiency in either human-rated or automatic evaluating systems. The American Council on the Teaching of Foreign Languages (ACTFL) oral proficiency test is a human-rated speaking test. In its guidelines, one of the major categories, 'Accuracy' is related to fluency.

> "The accuracy with which the task are performed. Factors included in this category include those traditionally in the foreground of oral language assessment: grammar, vocabulary, and pronunciation. In addition, OPI assessment recognizes the importance of fluency rate of delivery and coherence of message..." (Buck, 1999, p.8)

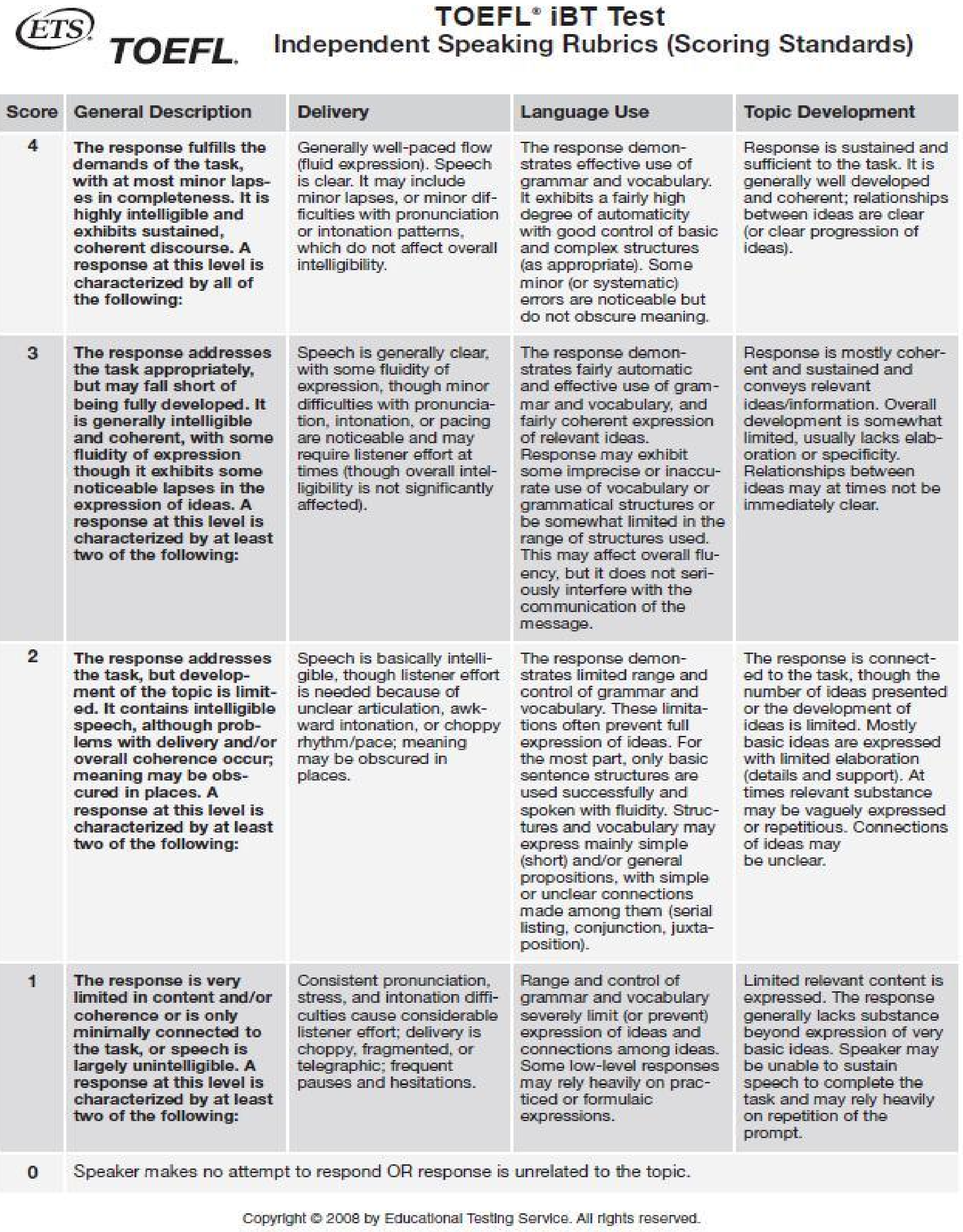
Further, the define fluency as the "rate of speech and the use of cohesive devices to bind discourse together (Buck, 1999, p.25)." In their analysis, there are four major levels: novice, intermediate, advance, and superior. In each major levels, except superior, there are three sublevels: high, mid and low.

- The "High" sublevel: Speakers at "High" sublevels communicate with ease..., but they are unable to sustain language at that next higher level without intermittent lapses or evidences of difficulty... The result is a lessening of the the overall performance, i.e., a form of linguistic breakdown that is generally mild and limited in scope: a drop of fluency... Advance-High speakers show solid and sustained ability in providing lengthy narrations...(Buck, 1999, p.18).
- The "Low" level: Speakers at the "Low" level summon up all their linguistic energy to sustain the requirements of the level. They exhibit less fluency and accuracy... (Buck, 1999, p.19)

From the ACTFL description of fluency, fluency is related to speaking rate and it reflects whether speakers are comfortable at the speaking task and can communicate effortlessly. Consider a testee who can speak fluently at an intermediate level. When he/she answers questions in a higher level, like the advanced level, he/she starts having frequent pauses, silence or repetitions. This is a sign of a drop of fluency, representing a linguistic break. This indicates that the testee can not sustain a higher level.

In another testing system - the Test of English as a Foreign Language (TOEFL) iBT speaking test, which is a computer-based test system. Fluency on the TOEFL iBT test is rated on a scale of 1 to 4. Fluency at the lowest proficiency level (score 1) is best described in the category of delivery as "choppy, fragmented, or telegraphic; frequent pauses and hesitations." At the next level (score 2), fluency is described as having a "choppy rhythm/pace" of speech. At a proficiency score of 3, a testee's "speech is generally clear, with some fluidity of expression." At the highest level (score 4), a testee's speech has a "generally well-paced flow (fluid expression)."

The complete rubric of the TOEFL iBT speaking test is shown in Figure 2.1 (*TOEFL iBT Test: Independent Speaking Rubrics (Scoring Standards)*, 2008). In both of the ACTFL and TPEFL, fluency is a crucial feature in language testing to evaluate oral proficiency.



ETS TOEFL

TOEFL® iBT Test
Independent Speaking Rubrics (Scoring Standards)

Score	General Description	Delivery	Language Use	Topic Development
4	**The response fulfills the demands of the task, with at most minor lapses in completeness. It is highly intelligible and exhibits sustained, coherent discourse. A response at this level is characterized by all of the following:**	Generally well-paced flow (fluid expression). Speech is clear. It may include minor lapses, or minor difficulties with pronunciation or intonation patterns, which do not affect overall intelligibility.	The response demonstrates effective use of grammar and vocabulary. It exhibits a fairly high degree of automaticity with good control of basic and complex structures (as appropriate). Some minor (or systematic) errors are noticeable but do not obscure meaning.	Response is sustained and sufficient to the task. It is generally well developed and coherent; relationships between ideas are clear (or clear progression of ideas).
3	**The response addresses the task appropriately, but may fall short of being fully developed. It is generally intelligible and coherent, with some fluidity of expression though it exhibits some noticeable lapses in the expression of ideas. A response at this level is characterized by at least two of the following:**	Speech is generally clear, with some fluidity of expression, though minor difficulties with pronunciation, intonation, or pacing are noticeable and may require listener effort at times (though overall intelligibility is not significantly affected).	The response demonstrates fairly automatic and effective use of grammar and vocabulary, and fairly coherent expression of relevant ideas. Response may exhibit some imprecise or inaccurate use of vocabulary or grammatical structures or be somewhat limited in the range of structures used. This may affect overall fluency, but it does not seriously interfere with the communication of the message.	Response is mostly coherent and sustained and conveys relevant ideas/information. Overall development is somewhat limited, usually lacks elaboration or specificity. Relationships between ideas may at times not be immediately clear.
2	**The response addresses the task, but development of the topic is limited. It contains intelligible speech, although problems with delivery and/or overall coherence occur; meaning may be obscured in places. A response at this level is characterized by at least two of the following:**	Speech is basically intelligible, though listener effort is needed because of unclear articulation, awkward intonation, or choppy rhythm/pace; meaning may be obscured in places.	The response demonstrates limited range and control of grammar and vocabulary. These limitations often prevent full expression of ideas. For the most part, only basic sentence structures are used successfully and spoken with fluidity. Structures and vocabulary may express mainly simple (short) and/or general propositions, with simple or unclear connections made among them (serial listing, conjunction, juxtaposition).	The response is connected to the task, though the number of ideas presented or the development of ideas is limited. Mostly basic ideas are expressed with limited elaboration (details and support). At times relevant substance may be vaguely expressed or repetitious. Connections of ideas may be unclear.
1	**The response is very limited in content and/or coherence or is only minimally connected to the task, or speech is largely unintelligible. A response at this level is characterized by at least two of the following:**	Consistent pronunciation, stress, and intonation difficulties cause considerable listener effort; delivery is choppy, fragmented, or telegraphic; frequent pauses and hesitations.	Range and control of grammar and vocabulary severely limit (or prevent) expression of ideas and connections among ideas. Some low-level responses may rely heavily on practiced or formulaic expressions.	Limited relevant content is expressed. The response generally lacks substance beyond expression of very basic ideas. Speaker may be unable to sustain speech to complete the task and may rely heavily on repetition of the prompt.
0	Speaker makes no attempt to respond OR response is unrelated to the topic.			

Figure 2.1: TOEFL iBT Test - Independent Speaking Rubrics

2.2.2 Fluency and Speech Planning

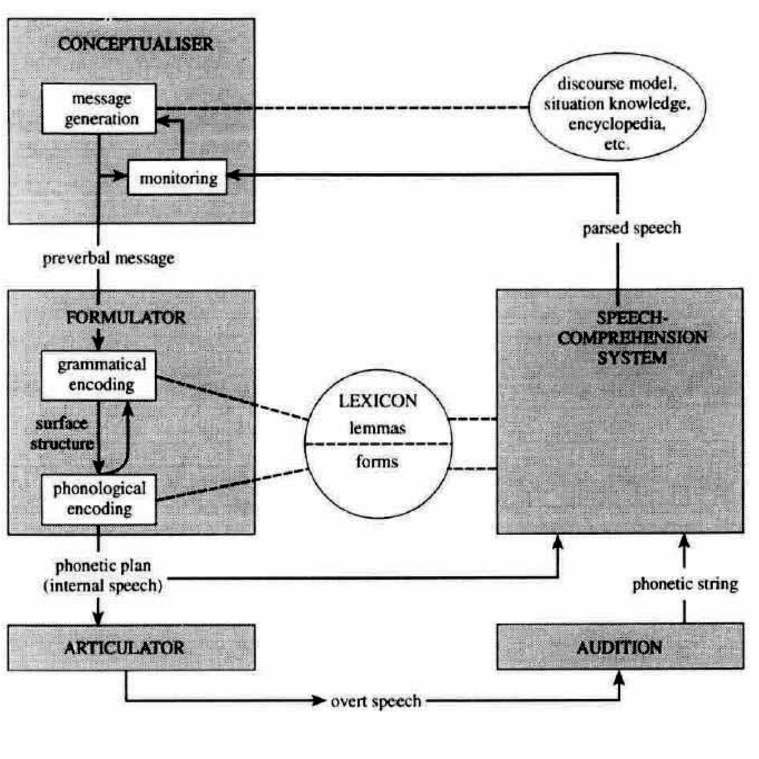


Figure 2.2: Levelt's model of language production (Levelt, 1989, p. 9).

What causes a lack of fluency? Levelt's model (Levelt, 1989) has great influence on contemporary ideas of speech production and it postulated a model for the generation of disfluencies. Figure 2.2 schematizes Levelt's representation of the mental process for the creation of fluent speech in L1 speech production. There are three main elements on the production (left side): conceptualizer, formulator and articulator. In the conceptualizer, the content of the intended message is developed as a form of 'preverbal message' by speakers. The output of the conceptualizer then is sent to the formulator, where the appropriate meaning and forms of lexical items are extracted to express the message. The grammatical encoding applies to form the surface syntactic structure and passes the message to the phonological encoding. A phonetic plan is delivered to the articulator to produce overt speech. Thus, fluency reflects multiple rapid and efficient processes in speech planning, including lexical access, grammatical and phonological encoding and the formation of speech into articulatory output.

The relationship between L1 fluency and L2 fluency has interested scholars for a long time. One view is that L2 fluency is independent from L1 fluency, whereas the opposite view suggests that L1 and L2 fluency are related. Derwing and Munro (2009) explored the relationship between L1 fluency and L2 fluency by examining temporal features in speakers' L1 and L2 speech production as well as human-rated fluency scores. Two groups of beginner-level English language learners, Slavic-speaking (Russian and Ukrainian) and Mandarin-speaking adults participated in their longitudinal study. Participants L1 productions were recorded at the beginning of their study, while L2 production was recorded across three data collection times, from the start of the study: after 2 months, after 1 year, and after 2 years. The fluency ratings revealed significant positive correlation between L1 and L2 productions recorded at 2 months from the start, but no significant correlation after 1 and 2 years later. As for the temporal measures, pause per second, speaking rate and pruned syllable per second had significant correlation between L1 and L2 productions at the data collection time of 2 months from the start in both speaker groups, while those measures presented significant correlation at times of 1 and 2 years later only in the Slavic group. The findings suggested that a variety of factors contribute to fluency development, including L1 structure (Russian vs. Mandarin), learners proficiency levels, the amount of exposure to L2 and so on. Thus, the relationship between L1 and L2 fluency is complex.

However, fluency is not the opposite of disfluency. Fluent speech still contains disfluencies, such as repairs, hesitations, and repetitions. In natural speech produced by native speakers, disfluencies affect up to 10% of the words and an overall 1/3 of the utterances (Shriberg, 2001). If a speaker keeps talking without any pauses, silence or breaks, it is hard for listeners to digest the messages conveyed. When a speaker is formulating speech and talking simultaneously, slowing down their speaking rate or pausing are strategies used to allow him/her to hold the speech floor or find the next focus (Chafe, 1980). In this sense, filled pauses (FPs), such as 'uh' and 'um' are viewed as "fillers" with similar functions as discourse markers 'you know' and 'well'. It has been proposed that 'uh' and 'um' are English words, which are planned for, formulated, and produced as parts of utterances just as other words are (Clark & Fox Tree, 2002). Rhus, even native speakers are unable to produce the required words sometimes and display problems in the process of producing speech, such as sttuttering and disfluencies (Goffiman, 1981; Levelt, 1989). Therefore, disfluencies in native speech may aid speech processing on the auditory side of listeners or impede the flow of speech on the articulatory side of speakers. This observation raises the question of what actually distinguishes native and non-native speech.

2.2.3 Quantitative Measurements of Fluency

How do we measure fluency quantitatively? Several studies in SLA have examined the correlation between objective measures and global fluency ratings by trained raters or untrained native L2 speakers. Möhle (1984) proposed various temporal features to measure fluency, such as speech rate(words/syllables per minute), length and positioning of silent pauses, length of fluent speech runs between pauses, frequency and distribution of filled pauses (FPs), repetitions, and self-corrections. He suggested that the difference between natives and non-natives lay in the frequency and distribution of those features rather than their presence in speech. Lenno (1990) investigated the performance of oral fluency produced by four L2 learners of English over time using a quantitative analysis. He observed that fluency scores improved during the second test after 21 weeks. He also observed that the speaking rate increased and the frequency of FPs decreased in L2 production. In Lennon's study, speaking rate was defined in two ways — the number of words per minute (including/excluding self-corrected words) and the frequency of FPs was defined as the number of FPs per T-Unit (Hunt, 1970; Vorster, 1980), where the T-unit refers to a main clause and all its attendant subordinate clauses and non-clausal units. Riggenbach (1991) studied the speech of six Chinese learners of English by examining their fluency, which was rated by English instructors as "fluent" or "nonfluent" and combined her analysis with pragmatic and temporal features. She found that the nonfluent learners have significantly more unfilled pauses (silent intervals in speech) than fluent learners.

Cucchiarini et al. (2000) have demonstrated that it is possible to predict the fluency rating of L2 learners using automatically calculated temporal measurements of speech quality taken from their read speech, including speech rate, articulation rate, number and length of pauses, number of disfluencies, mean length of runs, and phonation time ratio (defined as total duration of speech with/without pauses). In their experiment, the read speech of 20 native and 60 non-native speakers of Dutch was evaluated by expert raters, including phoneticians, teachers of Dutch as a second language, and speech therapists. The variables used in their study can be divided into three groups: speech rate, frequency effects and pause duration, including silent pauses and FPs. The results of their study suggest several points. First, all the variables are strongly related with fluency ratings, with the exception of pause length. Second, for fluency, the frequency of pauses is more relevant than their length. That is, native and non-native speech differ from each other more with pause frequency than with pause length. Third, rate of speech appears to be the best predictor for fluency because it incorporates articulation rate and pause frequency.

Cucchiarini et al. (2002) have further explored the relationship between temporal measures and perceived fluency in spontaneous speech. Due to the fact that pauses are more frequent in spontaneous speech than in read speech, the variables not containing information about the pause frequency have almost no relationship with perceived fluency.

Kormos and Dénes (2004) investigated the oral fluency of 16 Hungarian L2 learners at two proficiency levels through the automatic extraction of temporal measures from their speech. Also, they included comments about students' performance from three experienced native and three non-native Hungarian-speaking teacher. They suggested that pace, defined as the number of stressed words per minute, was also a good indicator in addition to other temporal features, such as speech rate, the mean length of an utterance, and the phonation time ratio.

Yoon (2009) developed an automated scoring method for L2 fluency rating, based on a set of automatic extracted temporal features grouped into three types, namely, speed (e.g. articulation rate, rate of speech), smoothness (e.g. frequency of disfluency/hesitation, mean duration of disfluency/hesitation) and syntactic complexity (e.g. number of words per clause). The result was compared to human ratings and showed that syntactic complexity had the least correlation with fluency scores, whereas speed, especially rate of speech, had the highest correlation with fluency scores.

Bhat (2010) developed an automatic estimator of vocabulary size and explored the effect of vocabulary size and temporal measures on L2 fluency in spontaneous speech. Her results suggested that raters were more strongly influenced by temporal aspects of L2 speech when they judge L2 fluency. She found that the lexical measures of word tokens and word types were good predictors of fluency as they demonstrated lexical richness in speech production, implying that larger vocabulary size decreased disfluencies and in turn relates to fluency.

Although much work related to FPs in English has been done, comparatively fewer empirical studies in Mandarin spontaneous speech have been carried out (Tseng, 2003; C.-K. Lin, Tseng, & Lee, 2005; Zhao & Jurafsky, 2005; Tseng, 2006). From the aspect of discourse analysis, Huang (1999) has reported that demonstrative pronouns, *na*/*nage* ('that') and *ranhou* ('and then') can serve as connective or pause markings as lexical fillers in speech. When two utterances are loosely connected, *na* ('that') sometimes functions pragmatically as either a logical connective ('then') or as a simple connective ('and'). Another use of the demonstrative *na* and *nage* ('that') is like an FP when the speaker has difficulty retrieving and searching lexical items and is planning the syntactic structure of utterances. Usually, hesitation, lengthening, or pauses are accompanied with the demonstratives *na* and *nage* when they function as lexical fillers.

Tseng (2003, 2006) investigated several types of repairs and repetitions with durational and structural analysis in Mandarin spontaneous speech. Lee et al. (T.-L. Lee, He, Huang, Tseng, & Eklund, n.d.) examined 786 prolongations in terms of the position in a sentence, the part of speech, and the segmental and tonal types. They found that prolongation often occurs at word/phrasal-final or utterance-medial positions and that consonants are rarely prolonged in Mandarin. The functions of prolongations are mainly produced for hesitation as well as for emphasizing a discourse focus. Zhao and Jurafsky (2005) have reported a descriptive study of FPs in Mandarin. Their research was based on the data from LDC 98-HUM5 Mandarin corpus of telephone conversations, in which the FPs ‘uh’ and ‘um’ are hand-labeled. The result shows that Mandarin speakers intensively use demonstrative *zhege* ‘this’ and *nage* ‘that’ as major types of FPs. Compared with the occurrence of ‘um’ and ‘uh’ in the CallHome English corpus, ‘um’ occurs 7.15 times per 1000 words, and ‘uh’, 7.1 times per 1000 words. In the 98-HUM5 Mandarin corpus, ‘um’ appears 1.46 times per 1000 words and ‘uh’ appears 2.55 times per 1000 words. They observed that the occurrence of ‘uh’ and ‘um’ is not frequent in Mandarin spontaneous speech. Further, the difference between demonstratives *zhege/nage* and ‘uh’/‘um’ lies in their distribution in different syntactic contexts. Demonstratives are more frequently used in a nominal-searching environment, while ‘uh’ and ‘um’ are more likely to be used at clause-initial position.

Wu (2008b) trained a Classification and Regression Tree (Breiman, Friedman, Olshen, & Stone, 1984) to determine whether native and non-native spontaneous speech can be predicted on the basis of temporal measurements of FPs. The result showed that *rate of speech* appears to be the best predictor in identifying native (F-score: .891) and non-native speakers (F-score: .853). The threshold of rate of speech for separating native and non-native speech is 3.14 syllables per second. Combining all variables (rate of speech, normalized frequency of FPs, mean length of FPs, normalized duration of FPs) results in excellent performance in distinguishing the native and non-native speeches, respectively (F-score: .920, F-score: .853).

On the basis of previous studies discussed in this section, rate of speech has the best predictive power of fluency. Information about the FPs ‘uh’ and ‘um’ were also useful indicators of fluency. Below are summaries of the temporal features that have been used (Riggenbach, 1991; Vanderplank, 1993; Ramus, Nespor, & Mehler, 1999; Cucchiarini et al., 2000; Kormos & De´nes, 2004).

- **Rate of Speech**: The total number of syllables produced in a given speech / Total duration of speech including pauses
- **Articulation Rate**: The total number of syllables produced in a given speech / Total duration of speech excluding pauses
- **Phonation Time Ratio**: Total duration of speech without pauses / Total duration of speech including sentence-internal pauses (*100)
- **Number of Silent Pauses**: Number of pauses of no less than 0.2 s per minute
- **Total Duration of Pauses**: Total duration of all sentence-internal pauses of no less than 0.2 s
- **Mean Length of Pauses**: Mean length of all silent pauses of no less than 0.2 s
- **Number of FPs**: Number of uh, er, mm, etc. per minute
- **Number of Disfluencies**: Number of repetitions, restarts, repairs per minute
- **Pace**: The number of stressed words per minute
- **Space**: The proportion of stressed words to the total number of words
- **Percentage of vowel duration**: The proportion of vocalic intervals within the sentence, that is, the sum of vocalic intervals divided by the total duration of the sentence (*100)
- **Standard Deviation of Vowel Duration**: The standard deviation of the duartion of vocalic interclas within each sentences

2.3 Foreign Accent

This subsection reviews the definition of foreign accent that is used in the SLA literature and the models that explain the factors leading to the phenomenon of foreign-accented speech. I will also present literature related to the measurements of the perception of foreign accent.

2.3.1 Definition of Foreign Accent

The definition of foreign accent has been formulated in a number of SLA studies. For example:

> "Foreign-accented speech, for instance, can be defined as nonpathological speech produced by second language (L2) learners that differs in partially systematic ways from the speech characteristics of native speakers of a given dialect." (Munro, 1998, p. 139)

> "the notion of "accentedness," which one might define as the extent to which an L2 learner's speech is perceived to differ from native speaker (NS) norms..." (Munro & Derwing, 1998, p. 160)

> "an instance of foreign accent consists in a deviation from the generally accepted norm of pronunciation of a language that is reminiscent of another language, i.e. the speaker's native language. It has to be emphasized that such a deviation must be defined in terms of its perception by listeners who are native speakers of the respective language and not in terms of differences in articulation that may be instrumentally measurable." (Jilka, 2000, p. 9)

According to the last two definitions, foreign accent refers to the way that speech deviates from native norms in terms of perception more than production. In other words, foreign accent is perceived when the speech of the speaker is different from that of the listener in a way that is reminiscent of another language, the L1. The characteristics of accent consist of speech differences in the pronunciation of consonants and vowels and prosody such as stress, tones, and intonation. For instance, Chinese speakers usually pronounce English interdental consonant [ɵ] as [s], e.g., 'thank you' becomes 'sank you'. Japanese, when using English, usually have problems producing [r] because Japanese does not make a distinction between [r] and [l]. Italian speakers tend to add [h] to some vowel-initial words in English.

Although accuracy of pronunciation and accent are not independent, in language assessment, accent is not used explicitly as a criterion to evaluate oral proficiency. Instead, the accuracy of pronunciation is one of the rating features in testing. In the ACTFL oral proficiency guidelines, pronunciation is defined as the 'ability to reproduce segmental and suprasegmental (pitch, stress, intonation) features of the language (Buck, 1999, p. 25).' In the TOEFL iBT speaking test rubric, the category of delivery is mainly concerned about the goodness of pronunciation. For the lowest proficiency level (score 1), "consistent pronunciation, stress, and intonation difficulties causes considerable listener effort." For the second lowest level (score 2), the speaker has "unclear articulation, awkward intonation." At the next level (scores 3), "minor difficulties with pronunciation, intonation, or pacing are noticeable and may require listener effort at time (though overall intelligibility is not significantly affected)." With the highest oral proficiency level (score 4), the speaker may have "minor difficulties with pronunciation or intonation patterns, which do not affect overall intelligibility".

From the definition from SLA, foreign accent is the perceptual distance of the speech between speakers and listeners (Munro & Derwing, 1998; Jilka, 2000). It usually reflects on pronunciation (sounds), prosody (stress, tones), or intonation.

2.3.2 Models of Second Language Acquisition

Why do learners speak with a foreign accent and how do listeners recognize foreign accents are questions that intrigue researchers. Several SLA models have discussed the formation of foreign accent and make predictions about how learners produce speech. In the following sections I will review the following models, the Native Language Magnet Model (NLM) (Kuhl & Iverson, 1995), the Perceptual Assimilation Model (PLM) (Best, 1995) and the Speech Learning Model (SLM) (Flege, 1995b). Perception studies investigating objective variables related to foreign accent will be presented as well.

Native Language Magnet Model

The Native Language Magnet Model (NLM) (Kuhl & Iverson, 1995) argues that exposure to L1 early in life alters the perceived distance between sounds of L1 and L2 in the acoustic space. As a consequences of the alternations in L2 perception, a speaker perceives an L2 sound with a bias and that is the source of foreign accent in the L2 production.

The NLM was developed from the observations of infants' early acquisition of speech perception before they could speak. Infants are capable of hearing sound differences in any language, but the perceptual sensitivity to non-native sounds reduces in adults as they become more attuned to the acoustic space of their native language. The distortion of the perceptual space is illustrated by the magnetic effect of phonetic prototypes attracting or reducing perceptual distances depending on the accurateness of the instances in the same category, as shown in Figure 2.3

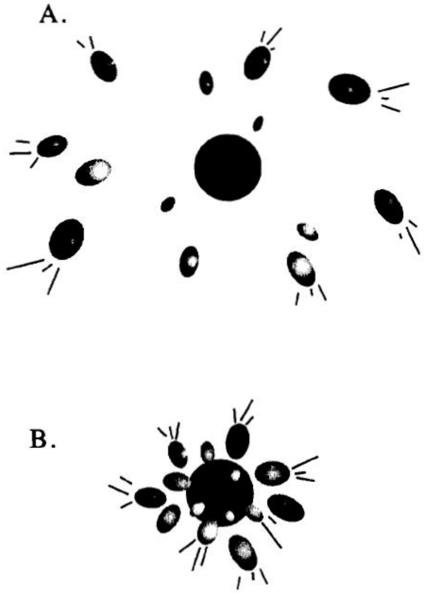


Figure 2.3: The perceptual magnet effect. Stimuli surrounding a phonetic prototype (A) are perceptually drawn toward the prototype (B), thereby shrinking the perceived distance between the prototype and other members of the category (Kuhl & Iverson, 1995, p. 124).

The existing L1 phonetic categories and perceptual boundaries shape the perception of L2 sounds and impede acquisition of non-native languages during adulthood. The magnet effect attracts L2 sounds which are similar to established L1 categories in the distorted perceptual space while L2 sound that are far away from the L1 prototypes enhance their discrimination from L1 sounds. A well known example is /r/ and /l/ in Japanese. These sounds are not distinguished by native Japanese speakers. Evidence from the discrimination of /r/ and /l/ by Japanese learners of English shows that /r/ and /l/ in English are attracted to the same Japanese categories and can not be distinguished by Japanese speakers (Iverson, 2003). Thus, Japanese speakers are unable to establish the perceptual categories of these two sounds separately. As a result, the lack of /r/ and /l/ distinction is often cited as contributing to a Japanese foreign accent.

The NLM model has drawn attention to issues of speech perception more than speech production. Since L2 sounds assimilated to L1 categories are not identical to those in the actual L2 acoustic space, production by L2 learners will lead to foreign accented speech. Some of the predictions that PAM makes are similar to the predictions that NLM makes. The difference between PAM and NLM is that PAM can be applied to production.

Perceptual Assimilation Model

The Perceptual Assimilation Model (PAM) (Best, 1995) addresses how L2 learners acquire L2 phones, incorporating articulatory phonology (Browman & Goldstein, 1986). The PAM assumes that sounds are perceived as articulatory gestures, which are organized hierarchically in a way similar to the shape of the vocal tract. Thus, articulatory gestures are the units of linguistic contrast of speech perception and production in the native phonological space.

In the PAM, non-native phones are perceived according to the similarities and differences of the articulatory gestures to the native sounds.

> “non-native segments, nonetheless, tend to be perceived according to their similarities to, and discrepancies from, the native segmental constellations that are in closest proximity to them in native phonological space. Because the universal phonetic domain and native phonological space are defined by the spatial layout of the vocal tract and they dynamic characteristics of articulatory gestures, those distal properties provide the dimensions within which similarity is judged.” (Best, 1995, p. 193)

There are three patterns of perceptual assimilation of non-native segments to native categories in the PAM (Best, 1995, p. 194).

1. If gesture similarities are detected, the non-native segments will be assimilated to a native category either as an acceptable (but not ideal) or a notably deviant exemplar of the category.
2. If gesture discrepancies are recognized, non-native segments will be assimilated as an uncategorizable speech sound within native phonological space (a new category), but one that does not belong to any native category.
3. If no similarity to any native category is detectable, a non-native segment is not assimilated into native phonological space, and is categorized as a nonspeech sound.

The predictions of the PAM have been tested widely on the perception of non-native phonemes but do not explicitly address the production of non-native speech in terms of foreign accent. It can be speculated that perceptual assimilation will affect the production of spoken language. Thus, the non-native segments assimilated to native categories might result in producing sounds with some degree of foreign accent, while production of the non-native sounds assimilated to new categories will not carry a foreign accent. For nonassimilable L2 sounds, the PAM predicts they are relatively easy for L2 acquisition. It may be difficult to learn the new articulation early on, but success in learning those sounds is predicted in the long run.

Speech Learning Model

The Speech Learning Model (SLM) (Flege, 1995b, 1995a, 2003) provides a theory of SLA for pronunciation that attempts to account for the segmental aspect (consonants and vowels) of a foreign accent. The SLM primarily investigates the role played by age-related factors in SLA. The general concept that L1 influences the L2, in terms of foreign accent and the relationship between production and perception, has been discussed in previous studies. The SLM elaborates on this concept and forms the model focusing on age-related limits on the ability to produce L2 sounds in native-like fashion. The SLM operates at the phonetic level of processing L2 pronunciation, not at the abstract phonemic level. The major learner group studied in this framework is bilingual speakers who have spoken their L2 for many years or over their life span.

In Flege (1995b), four postulates and seven hypotheses were proposed to explain the process of the acquisition of L2 pronunciation. According to the SLM, L2 learners need to detect the differences in sounds between an L1 and an L2 in order to establish new categories for the L2 sounds. However, such phonetic differences are not easy to discern if the onset age of learning is late, even when the length of residence in the L2 community increases. The basic idea of the SLM is that L2 sounds that are similar though not identical with L1 sounds are the most difficult to learn, because they are perceived to be similar. There are two mechanisms of classifying and processing L2 sounds in the SLM, namely. The first, phonetic category assimilation occurs when the category formation of L2 is blocked because some L2 sounds are too similar to L1 sounds and are identified as instances of L1 sounds (in other words, “equivalence classification”, whereby neighboring sounds in the L1 and the L2 may resemble one another). Following this hypothesis, similar L1 and L2 sounds are processed under a single category and the interaction between them is bi-directional, that is, L2 sounds might influence the L1 production as well. Second, phonetic category dissimilation occurs when a new category for the L2 speech sounds has been established so that the nearest L1 speech category deflects away from the L2 category in order to maintain contrast in the phonetic space. As a result, foreign accent is created due to the interference from the L1 but the L1 can also become accented due to the L2 influence. The reproduced L2 sounds are influenced by the L1 color which makes them distinct from the native sounds. If the L2 sounds are perceived as different, new categories will be established and produced with less degree of foreign accent.

In sum, the predictive power to capture foreign accent in each model focuses on different domain. The assumption of the NLM is based on speech perception and is difficult to be tested in speech production. The foundation of perception in the PAM is built on the articulatory gestures formulated in articulatory phonology, but not the actual articulation of sounds. A few studies have directly tested the PAM on real articulatory data. Most of the research related to PAM used a perception task and inferred from it that the differences in perception were due to articulatory differences. The SLM forms its prediction based on the phonetic domain of sounds. One of the crucial issue in the SLM models is that the definition of similarities between L1 and L2 sounds is not well-defined in these models. In the PAM, do we define sound similarities on the basis of articulatory features, such as the distinctive features in SPE (Chomsky & Halle, 1968) or feature geometry (Clements, 1985)? In the SLM, do we define the vowel similarities based on two domains (the first and second formants) of vowels produced by L1 and L2 native speakers or other features more than formant frequencies? Or, do we define similarities based on perceived distance between L1 and L2 phonetic categories? According to different methods of comparison, the vowel categories for predicting the L2 learnability will change and the predictions change.

In this thesis, the SLM will be adopted and evaluated to test its hypothesis in the acoustic data used in this study. The articulatory observation of Mandarin vowels will provide discussion related to the PAM predictions as well.

2.3.3 Perception of Foreign Accent

Adolescent and adult L2 learners usually carry some degree of foreign accent in their L2 speech production. Efficient communication between speakers and listeners relies on accurate acoustic structure and listeners' perceptions. Foreign accent is due to incomplete knowledge of L2 sounds and the transfer effect from L1. Both segmental (e.g. consonants and vowels) and suprasegmenatal (e,g. prosody, tones, intonations) aspects are affected and it happens at a phonemic, as well as a phonetic, levels. On the other hand, foreign accent is a perceptual phenomenon based on listener's judgements. Recent studies have shown that a listener's background, whether they have experience with foreign speech and their attitude, influence the perception of accented speech (Lindemann, 2000, 2005). Other research showed that listeners are able to adapt to foreign-accented speech over time and if they are exposed to foreign-accented speech from multiple talkers, rapid adaptation to the speech can be achieved (Bradlow & Bent, 2008). In this section, I will focus on reviewing the studies investigating features contributing to the perception of foreign accent.

A variety of studies have reported that several objective variables are related to the perception of accentedness, including segmental and prosodic errors, speaking rate, anf listener's background (Anderson-Hsieh, Johnson, & Koehler, 1992; Munro & Derwing, 1998, 2001; Munro, Derwing, & Morton, 2006; Lindemann, 2005). In SLA, a number of studies have investigated the relationship of perceived fluency, foreign accent and suprasegmental properties in L2 speech. Anderson-Hsieh et al. (1992) reported that the inaccurate L2 production of prosodic properties, such as stress, rhythm, and intonation might contribute to pronunciation ratings more strongly than inaccurate L2 segmental production. Trofovimich and Baker (2006) examined several suprasegmental features (stress timing, peak alignment, speech rate, pause frequency, and pause duration) in the production of adult Korean L2 learners of English and how each contributed to fluency and foreign accents. The findings show that the amount of L2 experience influenced the production of stress timing while the onset age of intensive L2 exposure influenced the other factors (speech rate, pause frequency, and pause duration). Moreover, only pause duration and speech rate appear to have significant predictive power of the perception of foreign accent.

L2 learners tend to speak slower than native speakers do. This can not be regarded as a transfer effect from L1 for not all L1 exhibit a slower speaking rate than L2. Instead, many factors cause L2 learners to speak at slower rates, including articulatory difficulties in segmental and prosodic accuracy, slower lexical retrieval, and incomplete syntactic and morphological knowledge (Munro & Derwing, 2001, p. 453). Flege (1988) examined the accentedness ratings after removing pauses from the sentences read by Mandarin and Taiwanese learners of English and reported no significant difference in accent ratings between the normal and modified conditions. The reason might be that the majority of removed pauses were too short in duration (about 200 ms). Munro and Derwing (1998) the manipulated speech rates of speech productions from 10 Mandarin learners of English to test its impact on listeners' judgements of accentedness. The original rate was compressed to a rate that was 10% faster and stretched to a rate that was 10% slower. They observed a curvilinear relationship between speaking rates and accentedness/comprehensibility ratings, as shown in Figure 2.4 and Figure 2.5. That is, both very fast and very slow speaking rates increase accentedness and decrease comprehensibility. Why did the fastest or the slowest speech receive lower accent ratings by listeners? It is not just because these speech samples were more severely modified. Munro and Derwing (1998) explained that very fast speech required extra processing and very slow speech retained information in short-term memory for longer time. Both of them caused processing difficulty and resulted in lower accent ratings. They found that the optimal speaking rate in L2 speech was a rate that was slightly faster than what L2 speakers usually used but one that was still slower than the typical speaking rates of native speakers. Although the definition of optimal speaking rates for L2 learners is not clear, this suggested that speaking rate influences foreign accent in addition to L1 transfer.

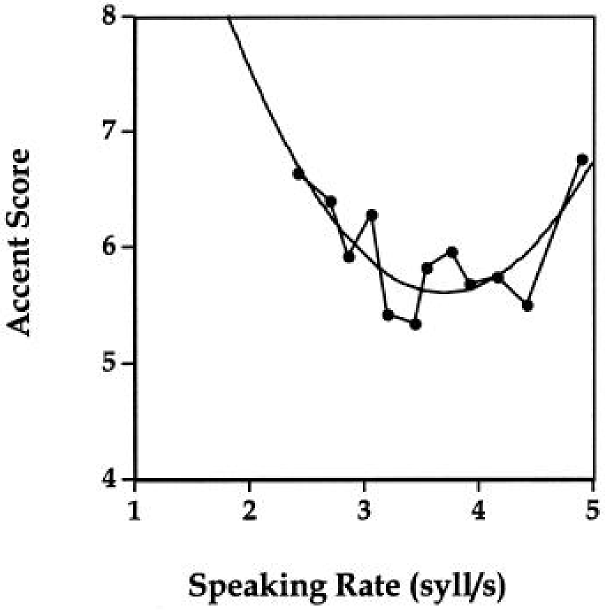


Figure 2.4: The relationship between speaking rate and judgments of accent (Lower score indicates less accented) (Munro & Derwing, 2001, p. 464).

Munro (1993) studied the relationship between acoustic measures of vowels and accentedness ratings by examining ten English vowels in /bVt/ and /bVd/

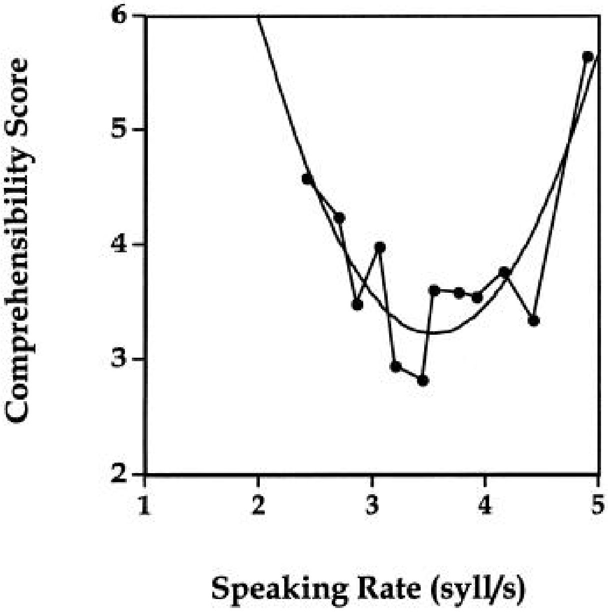


Figure 2.5: The relationship between speaking rate and judgements of comprehensibility (Lower score indicate higher comprehensibility) (Munro & Derwing, 2001, p. 465).

contexts produced by Arabic learners of English. The acoustic properties of vowels included vowel duration, F1, F2, and movement from F1 and F2. The findings reveal that accentedness ratings correlated with F1, followed by the degree of F2 movement. Further, different vowels show different magnitudes of difficulty for the L2 learners.

2.4 Summary

People with experience in foreign-accented speech often form an impression on what fluency and foreign accent are without well-defined concepts. In literature from language testing, speech production and speech perception, researchers attempt to provide definitions of fluency and foreign accent and make these notions concrete.

A shared presumption in recent theoretical models is that learners' abilities to discriminate L2 sounds is systematically influenced by their native sound systems. Speech perception becomes attuned to the native sound system during L1 acquisition and this increases the difficulty for a speaker to pick up non-native categories later on. The connection between perception and production explained by the models of SLA assumes that the L1 sounds are attractors, which is the source of foreign accent. However, the models make different claims in the level of assimilation and discrimination between L1 and L2 sounds. The PAM represents second language acquisition of sound systems at the abstract phonological space, while the SLM represents the processing of L2 pronunciation at the phonetic level (with age-related factors). The NLM does not explicitly address whether the sound prototypes are stored as abstract representations or as most representative instances of sounds surrounded by poor instances of the same category.

The purpose of this thesis is to clarify the concepts with precise definitions that can be tested experimentally and implemented technologically. This study investigated the topics of fluency and foreign accent by examining the relationship between human-rated judgements and acoustic measure of L2 speech, in order to figure out what acoustic attributes contribute to the perception of fluency and foreign accent. Is fluency only related to the acoustic measures, speaking rate and pauses? Learner's Mandarin vowel production will be studied to further inspect the contribution of pronunciation to foreign accent. According to the SLM, L2 learners need to detect the differences of sounds between L1 and L2 in order to categorize and establish L2 categories. It is hypothesized that L2 sounds that are similar to L1 categories are difficult to learn, while new sounds will be acquired after some level of initial struggle.

Chapter 3 examines the articulatory properties of Mandarin vowel categories and compares the acoustic similarities of Mandarin and English vowels to test the prediction of the SLM model. Chapter 4 studies spontaneous speech from corpus data that include heritage and English learners of Mandarin and native speakers. This study attempts to advance our knowledge of second language fluency and foreign accent in order to formulate educational and pedagogical strategies to help both learners and listeners.

Chapter 3

Mandarin and English Vowel Systems

In this chapter, I will introduce the phonetic and phonological categories of vowels in Mandarin first. Then, the pronunciation problem related to transliteration systems will be addressed. The literature regarding the assessment of segmental similarities across languages is reviewed. At the end of this chapter, my study will demonstrate the articulatory properties of Mandarin vowels. Because the predictions of SLM theories are based on the similarities between L1 and L2 sounds, the acoustic features between Mandarin and English vowels are compared.

3.1 Introduction

This chapter investigates Mandarin and English vowel systems in detail in order to evaluate how the English vowel system may affect a language learner's production of Mandarin vowels. Sound systems, such as consonants, vowels, tones, and intonations influence L2 production. The aquisition of tones in particular causes a lot of problems for L2 learners (Shih & Lu, 2010). In contrast to the inventory of consonants and tones existing in Mandarin, there is no consensus on the inventory of vowels in Mandarin and therefore, Mandarin-English vowel pairs are of particular interest. Thus, this thesis will only focus on vowels.

The Mandarin vowels we study in this thesis include phonemic vowels and their allophonic variations. Coarticulation affects the vowel quality greatly in Mandarin (Shih, 1995). In comparison, the analysis of English monophthongs in Peterson and Barney (1952) or other studies (Hillenbrand, Getty, Clark, & Wheeler, 1995; Hasegawa-Johnson, Pizza, Alwan, Cha, & Haker, 2003) is relatively clear, because the phonemic systems reflects the phonetic system. Thus, when learning Mandarin, English-speaking adult learners may encounter great difficulty learning some non-native vowels or even native-like vowels with non-native variations. Thus, the phonetic similarities and differences of vowels in Mandarin and English will be compared.

Mandarin learners in the US are frequently taught with the Pinyin system, which is a romanization system used to transliterate Mandarin sounds and provides a study aid to pronunciation. However, both of these factors (i.e., the fact that the Pinyin system does not reflect the vowel quality in coarticulation, and the fact that the mapping is inconsistent) can lead L2 learners to make errors in their productions. The spelling of some English words which are identical to the Pinyin spelling of Mandarin words is another plausible factor leading to accented speech. All of these issues will be discussed.

3.2 Mandarin Sound System

3.2.1 The Phonetic and Phonological System of Mandarin

Mandarin Chinese has relatively simple syllable structures (Chao, 1968; Shih & Ao, 1997; Duanmu, 2000; Y.-H. Lin, 2007). The number of segments in Mandarin syllables ranges from one segment, V, to a maximum of four, CGVC, where

Syllable Structure	Words in IPA	Gloss
V	i	‘one’
CV	mɑ	‘mother’
VC	an	‘safe’
GV	jɑ	‘a duck’
VG	ɑw	‘concave’
CVC	san	‘three’
CVG	pɑj	‘to pat’
CGV	ɕjɑ	‘shrimp’
GVC	wan	‘a bay’
GVG	wej	‘authoritative’
CGVC	twan	‘to hold’
CGVG	ɕɥej	‘place name’

Table 3.1: Mandarin syllable structures

C is a consonant, V is a vowel, and G is a glide. The syllable structure V is composed with only one monophthong. CV, VC, GV, and VG are the syllable structures with two segments. CVC, CVG, CGV, GVC and GVG are the syllable structures with three segments, where only the alveolar nasal /n/ or the velar nasal /ŋ/ can appear as the C after V in CVC, and only /j/ or /w/ can occur as the G after V. The maximum syllable structures are CGVC and CGVG. Table 3.1 gives examples of words with tone 1 in each syllable structure.

There are four contrastive tones plus a neutral tone in Mandarin (Chao, 1968; Cheng, 1968, 1973; Shih, 1986). Tone 1 is high level tone; Tone 2 is a rising tone; Tone 3 is a falling-rising tone, which has a low-falling shape with a rising tail; and Tone 4 is a falling tone. There is no controversy of the number of tones in Mandarin, where there are five phonemic tonal distinctions, including four lexical tones and the possibility of a syllable not carrying a tone phonologically. Mandarin also has tone sandhi rules, which are phonological rules that shuffle some tonal categories. After the application of these rules, however, there are still the same five tonal distinctions as the output of the phonological process (Shih, 1986; Speer, Shih, & Slowiaczek, 1989; Shih, 2008).

There are 22 consonants in Mandarin (Chao, 1968; Cheng, 1973; Ramsey, 1987). All of the stops and affricates are voiceless with an aspirated and non-aspirated contrast. The post-alveolar fricatives and affricates are retroflex and only the retroflex fricatives have a voicing contrast. Mandarin has three glides, labiovelar [w], palatal [j], and the front rounded palatal glide [ɥ]. Table 3.2 lists the phonetic inventories of consonants, glides and vowels. The shaded area indicates voiced sounds.

	BILABIAL		LABIL-DENTAL		ALVEOLAR -DENTAL		ALVEOLAR		POST-ALVEOLAR		PALATAL		VELAR		GLOTTAL
PLOSIVE	p pʰ				t tʰ								k kʰ		
FRICATIVE			f				s		ʂ	ʐ	ɕ				h
AFFRICATES							ts tsʰ		tʂ tʂʰ		tɕ tɕʰ				
NASAL		m				n								ŋ	
LATERAL						l									
GLIDE		w										j ɥ			

【VOWELS】

	FRONT	CENTRAL	BACK
HIGH	i y	ɨ	ɯ u
MID	e ɛ	ə ɚ	o ɔ
LOW	a		ɑ

Table 3.2: The sound system in Mandarin

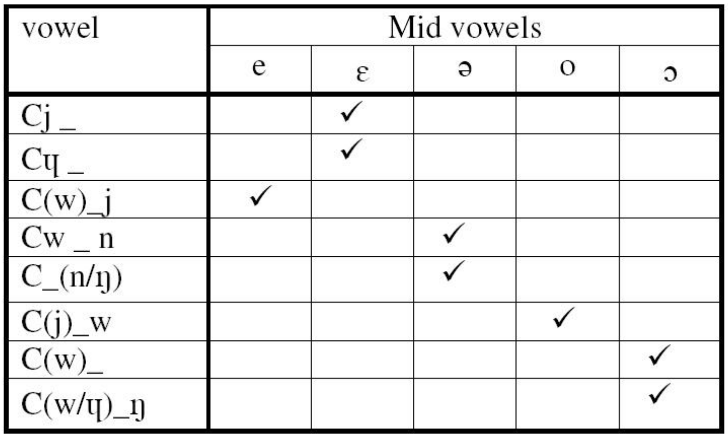
Phonologically, there are 13 monophthongs in Mandarin, including a retroflex vowel. These vowels are classified in three dimensions, namely: [high/low], [front/back], and [rounded/unrounded]. There are five high vowels, [i, u, y, ɨ, ɯ], while [ɨ] and [ɯ] only occur in CV syllables with alveolar and retroflex sibilants, respectively. Some scholars view [ɨ] and [ɯ] as voiced extensions of the preceding consonants and call them apical vowels (Chao, 1968; Cheng, 1973; Duanmu, 2000). These two high vowels and the preceding consonants are homo-organic. According to Ladefoged and Maddieson (1996), these apical vowels occur after affricates or fricatives and carry over the articulatory position of frication from the preceding consonants. Thus, they are also called fricative vowels.

> "Fricative vowels can usually be thought of as syllabic fricatives that are allophones of vowels. The best know examples are the allophones of i that occur after retroflex (Flat post-alveolar) and alveolar fricatives and affricates respectively in Standard Chinese. These vowels are made with the tongue in essentially the same position as in the corresponding fricatives. Because of the articulation used in the alveolar case, these vowels have sometimes been referred to as 'apical' vowles. This term is not appropriate for the so-called retroflex cases." (Ladefoged & Maddieson, 1996, p.314)

vowel	High vowels				
	i	ɨ	ɯ	y	u
p, p^h, m	✓				✓
f					✓
t, t^h	✓				✓
n, l	✓			✓	✓
s, ts, ts^h		✓			✓
ʂ, tʂ, tʂ^h,ʐ			✓		✓
ɕ, tɕ, tɕ^h	✓			✓	
k, k^h, x					✓

Table 3.3: Occurrence of consonants and high vowels

Lee and Zee (2003, p. 111) treated [ɨ] and [ɯ] as one sound. When following alveolar sibilants, the vowel is realized as a syllabic apico-laminal or laminal denti-alveolar approximant; when following post-alveolar sibilants, it is a syllabic apical post-alveolar approximant. Table 3.3 presents the co-occurrence constraints on consonants and high vowels in the syllable structure CV. The distribution of [y] is much more restricted than other vowels. /i, y, u/ are phonemes and they occur after [n] or [l]. The disagreement in the phonological analysis of high vowels lies in whether [ɨ] and [ɯ] are considered as the allophones of [i] or as an unspecified vowel.



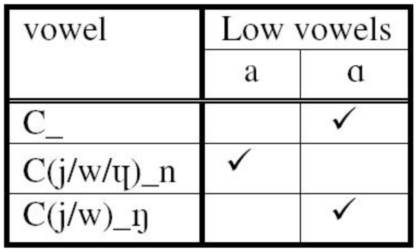
vowel	Mid vowels				
	e	ɛ	ə	o	ɔ
Cj _		✓			
Cɥ _		✓			
C(w)_j	✓				
Cw _ n			✓		
C_(n/ŋ)			✓		
C(j)_w				✓	
C(w)_					✓
C(w/ɥ)_ŋ					✓

Table 3.4: Occurrence of consonants and mid vowels

Mandarin mid vowels have several variants and the transcription of mid vowels in the phonological output is not consistent in the literature. In general, scholars agree that phonological outputs of mid vowels are generated from either /ɛ/ or /ə/ (Cheng, 1973; Y.-H. Lin, 1989; Wang, 1993; Duanmu, 2000). Through the assimilation process, there may be up to five surface variants, such as [e, ɛ, ə, o, ɔ]. Table 3.4 shows the distribution of mid vowels in different contexts (a parenthesis indicates an optional component). [e, ɛ, o, ɔ] can be analyzed as the allophones of

/ə/ as they occur in complementary distribution. The vowel [e] occurs before glide [j] (e.g. [gei3], *gei*, ‘give’) or in the context of Cwj (e.g. [gwej4], *gui*, ‘expensive’). Examples are given with IPA in squares, tones in numbers, Pinyin in italic fonts, and glosses in quotation marks. [ɛ] may only follow one of the glides [j, ɥ] in CGV syllables (e.g. [tje2], *die*, ‘to fall down’; [ɕɥɛ2], *xue*, ‘to learn’; the vowel [ə] exists in CV(N) (e.g. [kə1], *ge*, ‘a song’; [fən1] *fen*, ‘to distribute’) or CGV(N) syllables when G is the glide [w] (e.g. [twən1], *dun*, ‘squat’). The vowel [o] appears in C(j)VG, where G is the glide [w] (e.g. [tow1], *dou*, ‘all’; [tjow1], *diu*, ‘to throw’). [ɔ] presents in C(w)V or C(G)VN where G is glides [w, ɥ] and N is [ŋ] (e.g. [gwɔ1], *guo*, ‘wok’; [tɔŋ1], *dong*, ‘east’; [ɕɥɔŋ2], *xiong*, ‘bear’).

With regard to low vowels, the back low vowel [ɑ] occurs in an open syllable or before the velar nasal [ŋ]. The front low vowel [a] only occurs before the alveolar nasal [n], as shown in Table 3.5.



vowel	Low vowels	
	a	ɑ
C_		✓
C(j/w/ɥ)_n	✓	
C(j/w)_ŋ		✓

Table 3.5: Occurrence of consonants and low vowels

Cheng (1973) suggests the underlying low vowel in Mandarin is the low back vowel /ɑ/ that appears in open syllables. Since the alveolar nasal /n/ is front and the velar nasal /ŋ/ is back, the backness of the underlying low vowel is determined by the following consonants, as shown in the rewrite rule (1). Most scholars agree that there are two low vowels in the surface form.

1. ɑ –> a / _ n

Howie (1970) was the first study to define systematic acoustic properties of Mandarin vowels. He reported formant values of vowels, including phonemes and allophones with the occurrence of four tones. However, his work is more descriptive oriented and lacks further analysis. Shih (1995) investigates the acoustic properties of Mandarin vowels for the purpose of achieving naturalness in text-to-speech synthesis. The findings can be summarized as follows: (1) The coarticulation effects of the following nasals [n, ŋ] in [ən] and [əŋ] and the preceding glides [j, w] in [jɑ] and [wɑ] are expected and consistent with the anticipated tongue position of the sounds in context. (2) For diphthongs, the vowel nucleus is typically different from the corresponding monophthongs, with the exception of the diphthong [ow]. That is, [ɑ] in [ɑj], [ɑw] and [e] in [ej] are different from monophthongs [ɑ] and [e], respectively. Diphthongs are only similar to the corresponding monophthongs at the beginning 20% portion of vowel duration.

In sum, because of the patterns of complementary distribution, the five phonetic high vowels form three or four phonological categories. The five phonetic mid vowels are treated as one phonological category. The two low vowels form one phonological input. In this thesis, I will examine the 12 phonetic monophthongs in native and learners production for further analysis.

3.2.2 Articulatory Study of Mandarin Vowels

A few articulatory studies of Mandarin vowels are reviewed here. In an X-ray study by Wu and Lin (1989), the vowels [i, y, u, o, ɑ] were produced once by each of the study's five participants. The tongue, jaw, lip positions of the vowels [i, y, u, o, ɑ] were summarized as follows.

1. Vowel [i] has the highest tongue height followed by vowels [y], [u], [o], and [ɑ].
2. Vowel [ɑ] has the widest jaw opening followed by vowels [o], [u], [y], and [i].
3. Vowel [ɑ] has the largest lip aperture followed by vowels [i], [o], [u], and [y].
4. Only three participants demonstrated slight lip protrusion for the rounded vowels as compared to the unrounded vowels.

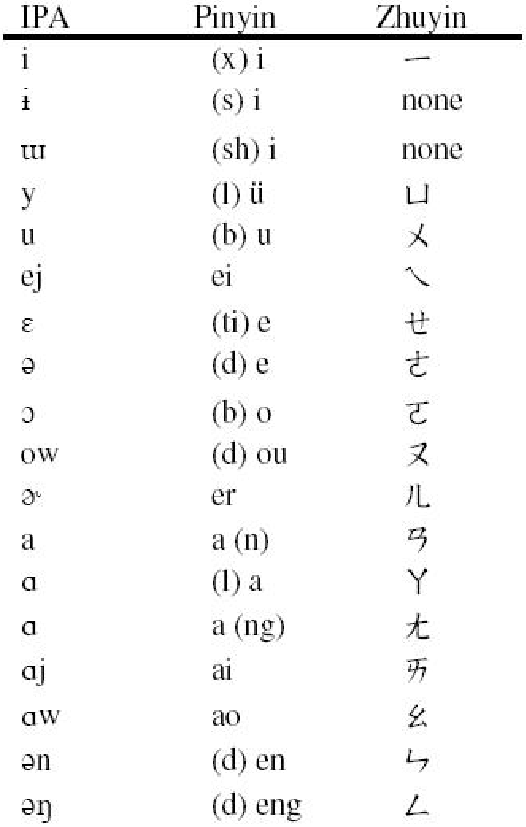
In an Electromagnetic Midsagittal Articulography (EMMA) study by Torng (2000), 24 words consisting of five Mandarin vowels [i, y, u, ɑ, o] with four tones were measured. The results are summarized as follows:

1. For the tongue body position, the high vowel [i] has the highest absolute tongue height followed by [y], [u], [o], [ɑ], as expected.
2. For the jaw position, vowels [u, y] have high jaw positions and the vowel [ɑ] has a low jaw position. Unexpectedly, the mid vowel [o] has a jaw position as high as the high vowel [u] and the high vowel [i] has a lower jaw position than [u, y, o].
3. Vowels [y, u, o] have stronger lower lip protrusion and vowels [i, ɑ] have weaker lip protrusion. As expected, vowels [y, u] have a smaller lip aperture and vowels [i, o, ɑ] have a larger lip aperture.

   The tongue body position is determined by the jaw position since the tongue rests on the jaw. One exception is the vowel [o]. The derived tongue position shows that the vowel /o/ has a lower tongue position than the vowel /ɑ/.

3.2.3 Transcription Systems and Pronunciation Errors

Another obstacle for the acquisition of vowels by L2 learners of Mandarin may relate to how the vowels are taught in the language classroom. The Chinese writing system uses symbols known as Chinese characters. This system is not an alphabetic writing system and the mapping from writing to sounds is not transparent. Thus, many transliteration systems were developed to annotate pronunciations as an educational aid in school systems. The transliteration systems



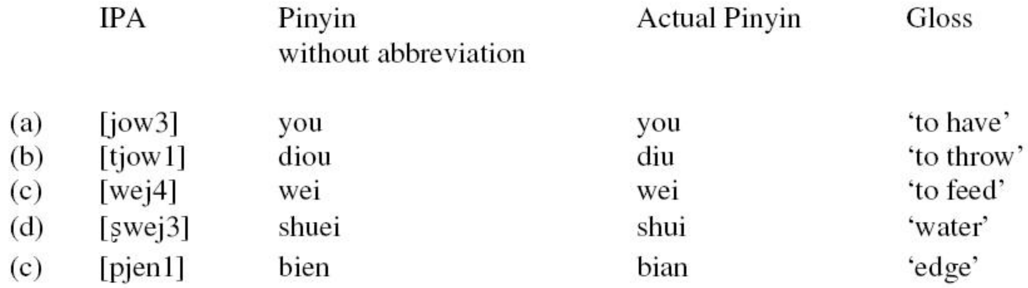
IPA	Pinyin	Zhuyin
i	(x) i	一
ɨ	(s) i	none
ɯ	(sh) i	none
y	(l) ü	ㄩ
u	(b) u	ㄨ
ej	ei	ㄟ
ɛ	(ti) e	ㄝ
ə	(d) e	ㄜ
ɔ	(b) o	ㄛ
ow	(d) ou	ㄡ
ɚ	er	ㄦ
a	a (n)	ㄢ
ɑ	(l) a	ㄚ
ɑ	a (ng)	ㄤ
ɑj	ai	ㄞ
ɑw	ao	ㄠ
ən	(d) en	ㄣ
əŋ	(d) eng	ㄥ

Table 3.6: Conversion chart of Zhuyin, Pinyin and IPA symbols

reflect the developer's views of the sound system and they may affect children and language learners' phonological development. Today, the two most commonly used systems are Zhuyin and Pinyin. Zhuyin is used in Taiwan and Pinyin is used in mainland China, as well as in most of the textbooks in the U.S. for learning Chinese as a foreign language. Zhuyin uses distinct symbols to represent 21 onset consonants, 3 glides, and 13 rhymes. Two of the rhymes ([ɨ, ɯ]) are not assigned distinct symbols, and the 3 glides are also used to represent vowels. Hence, this system recognizes a maximum of 18 rhymes, if the two vowels [ɨ, ɯ] are treated as two categories. Pinyin is a romanization system that uses Roman letters to present sounds in Mandarin. Vowels are indicated by the six symbols 'a, e, i, o, u, ü'. These two transcription systems give us the range of the possible number of surface vowels in Mandarin: anywhere between 6 and 18. Table 3.6 shows the correspondence of the annotations pertaining to Mandarin vowels in IPA, Pinyin and Zhuyin systems. In Pinyin, the letter *i* is used to indicate three high vowels [i, ɨ, ɯ]. The letter *e, o* indicated the mid vowels [e, ɛ, ə] and [o, ɔ], respectively. The low vowels [a, ɑ] are both displayed with the letter *a*. Additionally, Pinyin uses *a* to represent *e* in syllables like *tian* ([tjen], 'sky'). However, the vowel pronunciation indicated by the same letter may pronounce differently by native speakers. Moreover, many of the pronunciations in Mandarin are quite different from that of the pronunciation of the alphabets in English. As a result, the use of Pinyin may lead to L2 pronunciation problems.

Furthermore, Pinyin has additional conventions that are confusing to language learners. The first case is that sounds are sometimes spelled out and sometimes are not, i.e., some three-sound sequences are abbreviated and written only with two letters. As shown in Table 3.7 (a), 'you' is written as three letters, but the same sound is shortened as two letters in (b). The rhyme of the word *diu* is the same as 'you', but is written as 'iu', while the pronunciation is closer to 'diou'. The glide *y* and *w* is written as *i* and *u*, respectively when there is an onset consonant. Similarly, another example shows that 'wei' in Table 3.7 (c) is written as 'ui' in (d), while the pronunciation of 'shui' 'water' is closer to 'shuei'.

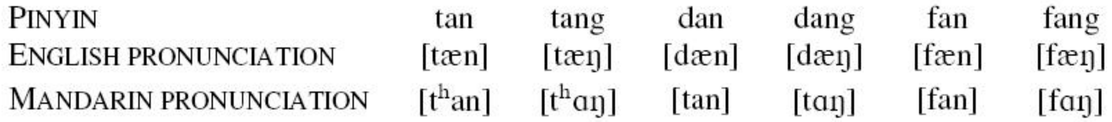
The second case of the Pinyin confusion is the pronunciation of 'ian' which sounds more like 'ien'. Cheng (1973) suggested that the low front vowel /a/ changes to the mid vowel [ɛ] if it is preceded by the high front vowel /i/ and followed by the alveolar nasal /n/. This suggests that the vowel in the sequence [ian] is [ɛ], instead of [a]. In Table 3.7 (e), the pronunciation of the word *bian* ‘edge’ in Pinyin is closer to ‘bien’. Based on these observations, it is postulated that transliteration confusion will contribute to difficulties in Mandarin vowel acquisition for language learners. It can potentially mislead them into acquiring a non-target-like representation of these sounds.



	IPA	Pinyin without abbreviation	Actual Pinyin	Gloss
(a)	[jow3]	you	you	‘to have’
(b)	[tjow1]	diou	diu	‘to throw’
(c)	[wej4]	wei	wei	‘to feed’
(d)	[ṣwej3]	shuei	shui	‘water’
(c)	[pjen1]	bien	bian	‘edge’

Table 3.7: Pinyin Confusion

Another problem of Pinyin is that the Roman letters do not differentiate phonetic categories pronounced differently by native speakers. One case is that the high vowels [i, ɨ, ɯ] are all indicated with the letter *i*, suggesting to language learners that the pronunciation is the same for these three sounds. This will lead to mispronunciation for language learners (Y.-H. Lin, 2007, p. 72). Another case is in regards to the low vowels [ɑ] and [a]. They are allophones and both are represented by the letter *a*, implying that the coarticulation effect to distinguish these two vowels are not introduced through Pinyin. Two issues are of concerned for these vowels. First, language learners might pronounce these two low vowels the same since the orthographic symbol does not differentiate them. Second, the correspondence between Pinyin and orthography of English words might confuse learners. In English, the vowels in the words in Table 3.8 are consistently the mid front vowel [æ], while in Mandarin, the vowel qualities are conditioned by the following codas, i.e., [a] followed by [n] and [ɑ] followed by [ŋ]. Phonetically, the [a] (e.g. *an*, ‘safe’) in Mandarin is different from the [æ] (e.g. Ann) in English. In terms of articulation, [æ] has lower tongue position accompanied by the spreading of the lower lips. Thus, the coarticulation effect of Mandarin low vowels as well as the orthography confusion between Mandarin and English result in difficulty in pronunciation acquisition for language learners.



PINYIN	tan	tang	dan	dang	fan	fang
ENGLISH PRONUNCIATION	[tæn]	[tæŋ]	[dæn]	[dæŋ]	[fæn]	[fæŋ]
MANDARIN PRONUNCIATION	[tʰan]	[tʰɑŋ]	[tan]	[tɑŋ]	[fan]	[fɑŋ]

Table 3.8: Orthographic Confusion of Pinyin and English

Wu (2008a) investigated the vowel space from spontaneous speech production by one female Mandarin native speaker and one female English-speaking L2 learner of Chinese in the classroom setting. Figure 3.1 presents the vowel formant chart, where the subscript 1 indicates native production and the subscript 2 indicates the learner’s production. It is observed that the vowel space of low vowels [a] and [ɑ] is flipped in L2 production. The L2 learner produced the low back vowel [ɑ] as being more fronted and higher, whereas the front low vowel [a] was pronounced as a back low vowel. The learner showed two vowel distinctions, suggesting that she learned the coarticulation effect in Mandarin low vowels, probably from the Chinese language classroom. However, the low back vowel [ɑ] is more fronted and similar to the English [æ]. In other words, the allophonic distribution of low vowels, the orthographic transcription, and the L1 sound system might all cause confusion for L2 learners and lead them to produce non-native like targets. This pattern was observed again in vowel productions by English learners in the corpus data, which will be reported in Chapter 4.

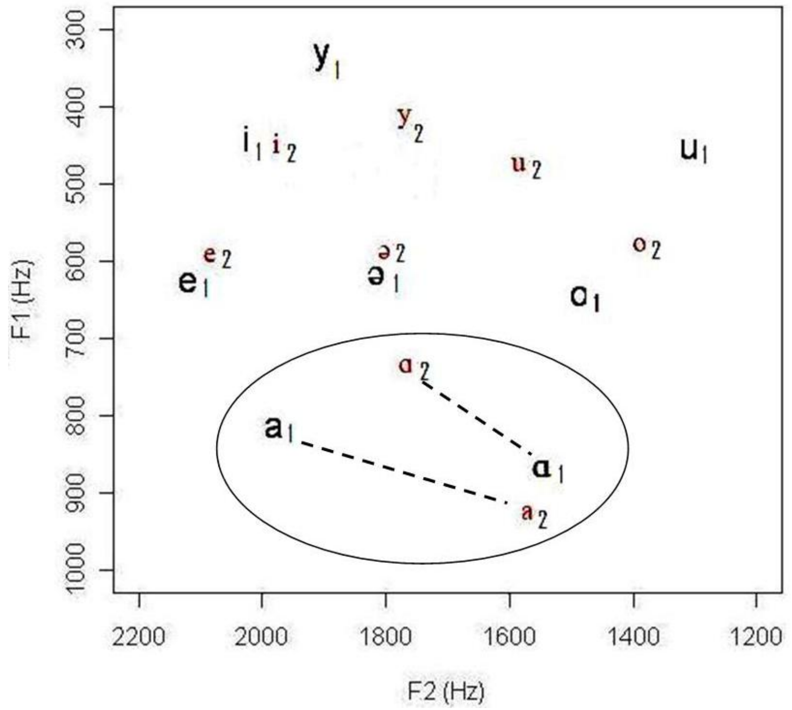


Figure 3.1: The vowel space produced by one Mandarin native speaker and one L2 learner. The subscript 1 indicates mean formant values of vowels produced by the native speaker and subscript 2 indicates the production by the L2 learner. (C.-H. Wu, 2008a)

3.3 Assessment of Segmental Similarities

Assessing similarities of speech sounds across languages is a crucial topic that needs to be addressed for the purpose of predicting L2 learner behavior in SLA models. Best's PAM (Best, 1995) addresses L2 learner's perception of non-native sounds based on how similar they are articulated with L1 sounds. The more closely the L1 and L2 sounds are articulated, the more likely they are to be assimilated. Non-assimilated L2 sounds are perceived as uncategorizable speech sounds. Flege's SLM predicts the ability of L2 sounds over a life span based on the interaction between L1 and L2 categories. 'Similar' sounds are perceived as instances of L1 categories; 'new' sounds are perceived as different from L1 categories and thus form a new category.

Numerous methods have been proposed to compare segmental categories across languages. Perceptual mapping is one of the approaches for measuring similarities. Guion et al. (2000) conducted an AXB discrimination task to test English consonant recognition by Japanese native listeners with varying amounts of English-language exposure. In the task, subjects heard three sounds in the sequence of A, X, and B and were asked to indicate whether the sound X is more similar to A or B. Best et al. (Best, McRoberts, & Goodell, 2001) used am AXB discrimination task to test the differences of sounds in perception and asked subjects for the transliteration of those sounds in the English orthography. In the study of Strange et al. (2004), the perceptual similarity of vowels in American English and North German were compared. Native English listeners were presented with key words, such as *heed, hid, head* etc. and the IPA symbols of each vowel category were listed beside the key words. Listeners heard the north German vowels and were asked to click on the English word which is most similar to the vowel that they heard. Then, listeners heard the same stimuli again and were asked to rate the accentedness of the stimuli on a 7-point scale from 'very foreign-sounding' to 'very native-sounding'.

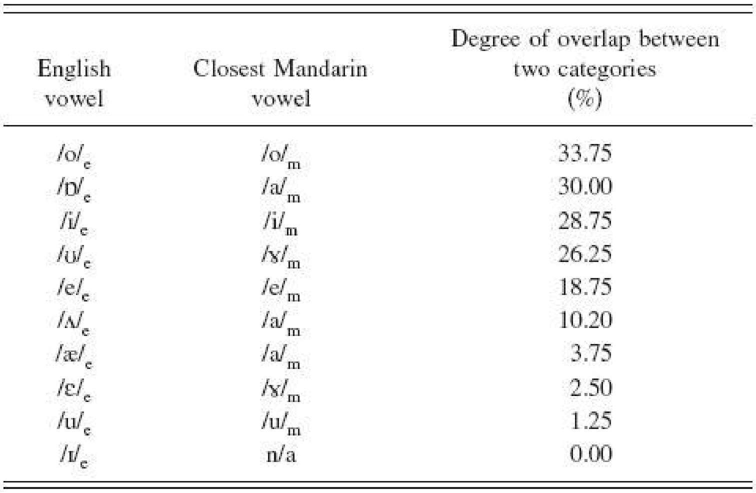
Other methods for evaluating cross linguistic similarities of vowels is to measure the extent of the spectral overlap of the vowel distribution. The spectral comparison usually takes the first (F1) and second (F2) formant values at the midpoints of each vowel duration. Bohn and Flege (1992) compared the spectral properties of the English /i, ɪ, ɛ/ with the German /i, ɪ, ɛ, ɛː/, produced by native speakers of each language, by plotting the formant space. Strange et al. (2004)

compared acoustic similarities of F1, F2, and F3 (the third formant) of vowels between North German and American English by examining spectral overlap.

Another method for assessing vowel similarities between L1 and L2 is to construct statistical pattern recognition models incorporating acoustic features, such as F1, F2, F3, pitch, duration, and spectral change (Strange et al., 2004; Morrison, 2006; Thomson, 2007; Thomson et al., 2009). This approach uses discriminant analysis to build a statistical model by using acoustic measurements of native vowel productions to define the L1 categories. Then, native productions of the L2 vowels are classified into the constructed L1 categories. Thomson et al. (2009) modified the single-language pattern recognition described above to train categories on both L1 and L2 languages. After training on all relevant categories in both languages, new cases from each language were tested to see how they are classified into the categories of the competing languages. Hence, the extent of misclassification provides a measure of assessing crosslinguistic similarity. Further, the modified model provides measurements of how well the new tokens fit into the other language category.

In the study of Thomson et al. (2009), vowel similarities between Mandarin /i, e, a, uə, u, o, ɤ/ and English /i, ɪ, e, ɛ, æ, ɑ, ʌ, o, ʊ, u/ were compared. All Mandarin and English vowels were elicited from /pV/ and /bV/ contexts. Acoustic vowel properties were extracted and submitted to pattern recognition models. These acoustic properties included F1, F2 and F3 sampled at 20% and 70% of each vowel duration, the mean F0 (fundamental frequency or pitch), and duration values of vowels Three pattern recognition models were trained: (1) a Mandarin pattern recognition model; (2) an English pattern recognition model; (3) a metamodel, testing Mandarin and English vowels within a single system.

Similarities were measured by averaging the degree of overlap between categories from each language. For instance, Mandarin /i/ is categorized as English /i/ 27.5% of the time, while English /i/ is classified as Mandarin /i/ 30% of the time (Thomson et al., 2009, p. 1452-1453). Thus, the similarity between Mandarin /i/ and English /i/ is the average overlap, 28.75%. Table 3.2 presents the degrees of overlap between English and Mandarin vowel categories. The results show that the most similar vowels are Mandarin /o/ and English /o/, followed by Mandarin /a/ and English /ɑ/, Mandarin /i/ and English /i/, Mandarin /ɤ/ and English /ʊ/ and to the least extent, Mandarin /e/ and English /e/. English /ʌ/ and /æ/ are less similar to Mandarin vowels. The model never classified English /ɪ, ɛ/ as Mandarin vowels and rarely misclassified English /u/ as a Mandarin vowel.



English vowel	Closest Mandarin vowel	Degree of overlap between two categories (%)
/o/$_e$	/o/$_m$	33.75
/ɒ/$_e$	/a/$_m$	30.00
/i/$_e$	/i/$_m$	28.75
/ʊ/$_e$	/ɤ/$_m$	26.25
/e/$_e$	/e/$_m$	18.75
/ʌ/$_e$	/a/$_m$	10.20
/æ/$_e$	/a/$_m$	3.75
/ɛ/$_e$	/ɤ/$_m$	2.50
/u/$_e$	/u/$_m$	1.25
/ɪ/$_e$	n/a	0.00

Figure 3.2: Degree of overlap between English and Mandarin vowel categories, ranked from highest to lowest. The subscript e indicates English vowels and the subscript m indicates Mandarin vowels (Thomson et al., 2009, p. 1453).

Thomson et al. also compare spectral overlaps of vowel distribution between Mandarin and English, as plotted in Figure 3.3. Similar with the result from

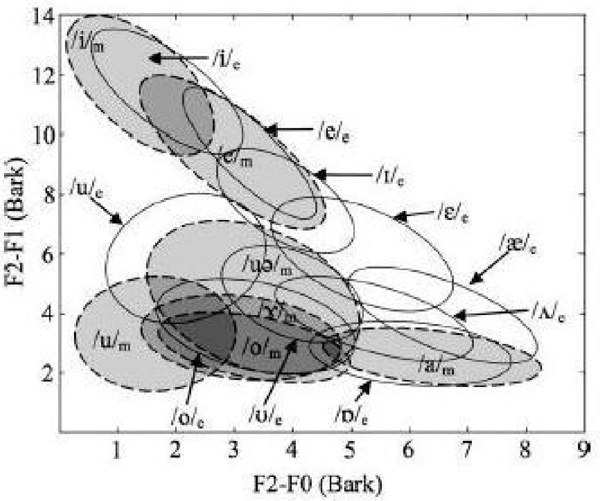


Figure 3.3: Ellipses representing Bark transformed midpoint values of L1 Mandarin (enclosed in broken lines with subscript m) and L1 English (enclosed in solid lines with subscript e) vowels (Thomson et al., 2009, p. 1453).

the model, the plot seemed to suggest that English /i, e, o ɑ/ are similar to Mandarin counterparts, while English /æ, ʌ/ are less similar. English /ɛ/ is quite different from Mandarin categories. Unlike the model's classifications, this plot suggests that English /ɪ/ and /u/ are somewhat similar to Mandarin /e/ and /uə/, respectively. It also suggests that English /ʊ/ is less similar to Mandarin /ɤ/. Since vowel categories overlap with one another within Mandarin and English, they argued that the information from the analysis of spectral overlap is less precise and difficult to quantify than the trained model.

In sum, the perceptual comparison reflects the similarities and differences of sounds across languages, but lacks information about production. The spectral assessment of similarities only relies on a few dimensions of the acoustic properties of vowels. The statistical analysis provides a better way to compare sounds crosslinguistically because it establishes the vowel categories in each language based on native productions and classifies new instances of each language into the competing languages. In addition, it incorporates multiple dimensions of acoustic attributes of vowels. However, none of the methods compares the similarity and discrepancy of sounds from real articulatory data. Further, if no agreement of sound similarities can be drawn among perception, acoustics, and articulation, which dimension should be used for comparison and to make predictions?

3.4 Comparing Vowel Systems in Mandarin and English

According to the SLA theories discussed in Chapter 2, assessing crosslinguistic vowel similarities is crucial for predicting and explaining L2 production. In order to compare Mandarin and English vowel systems and see how one affects the other in L2 learning, this section presents a study of the complete investigation of Mandarin and English vowels in terms of acoustics and articulation as a basis for the analysis of vowel production by L2 learners in the corpus study.

The data presented in this section was recorded using an Electromagnetic Articulography (EMA) AG500. Mandarin native speakers were asked to produce all possible Mandarin vowels. All of the Chinese learners speak English natively. They were asked to produced English vowels as well as Mandarin vowels. The articulatory data of Mandarin vowels that have been analyzed and the acoustic data for the comparison between Mandarin and English vowels will be presented in the following sections.

3.4.1 Method

Speakers

Mandarin vowel production data were obtained from 8 native speakers (4 females, 4 males; ages 20-30, M=25) from Taiwan. All were current or recent students at the University of Illinois at Urbana-Champaign. Their length of residence in the United States ranged from 7 months to 7 years (M=3 years). Their age of arrival was between 16 and 30 years of age (M=22.8 years). All reported normal hearing and speech.

The English vowel productions were obtained from 8 native speakers (4 females, 4 males, ages 19-30, M=24.3). The English native speakers were beginning learners of Chinese and had Chinese as a second language instruction between a few months to 2 years. All reported English as their first language and only language for daily communication, although they may have knowledge of several languages, such as Spanish and German. All had normal hearing and speech.

Mandarin and English Recording Materials

In the database, the Mandarin stimuli consisted of a set of approximately 400 possible syllables including all Mandarin vowels with tone 1 and a small set of syllables with the other three tones. This set included the syllables as shown in Table 3.9, with the syllables of the low vowel in context. Each syllable was read in the frame sentence, “Zhe ge ___ zi” meaning “This ___ word”, to avoid the lengthening effect of producing the test words in the final position of a sentence. The stimuli were repeated twice in a pseudo-random order. All stimuli were displayed in the written form of traditional Chinese characters and annotated with Zhuyin and Pinyin symbols.

Vowel	Pinyin
/i/	di
/ɨ/	si
/ɯ/	shi
/u/	du
/y/	lü
/e/	dei
/ɛ/	die
/ə/	de
/o/	dou
/ɔ/	bo
/a/	dan
/ɑ/	da / dang / jia / dao /dai

Table 3.9: Mandarin word list with all tones

Vowel	h_ d	b_ d(t)
/i/	heed	bead
/ɪ/	hid	bid
/u/	who'd	booed
/ʊ/	hood	-
/e/	hayed	bayed
/ɛ/	head	bet
/æ/	had	bad
/ʌ/	hud	bud
/o/	hoed	bode
/ɔ/	hawed	bawd
/ɑ/	hod	bod
/aɪ/	hide	bide
/aʊ/	howdy	bowed

Table 3.10: English word list

The English data contain all the English vowels in the contexts of /hVd/, and / bVd/, or /bVt/ (Peterson & Barney, 1952; Hillenbrand et al., 1995). All the stimuli were presented in a carrier phrase, “Say ___ again” and in pseudo-random order. The entire recording was repeated to obtain four repetitions of each item. Table 3.10 gives the English word list.

Procedure

L1 vowel productions of Mandarin and English speakers were recorded individually using an EMA AG500 in the Speech Dynamics Laboratory at the Beckman Institute at the University of Illinois at Urbana-Champaign. This apparatus consists of the EMA cube with six transmitter coils generating magnetic fields at different frequencies at defined orientations, with a receiver, 12 sensors, a computer with an automatic calibration unit, a real time display, and head movement correction systems. Sensors, which are built of small coils, are fixed onto the articulators of the subject. The alternating currents induced by the alternating magnetic fields have different strengths as a function of the distance and the angle of the sensor to the respective transmitter coil (*AG500 Manual*, n.d.).

A head-mounted microphone was worn by subjects and connected to the audio box, transferring the speech signal to the synchronizer. All articulatory and acoustic data are acquired and synchronized simultaneously through the EMA system and stored on the hard disk of the computer. Figure 3.4 is a schematic drawing of the setup of the system.

The procedure includes system calibration, sensor placement, and data collection. The EMA system needs to be turned on for at least one hour to reach a stable temperature. The system also needs to be calibrated with the EMA automatic calibration system in order to minimize the mean spatial error for the data.

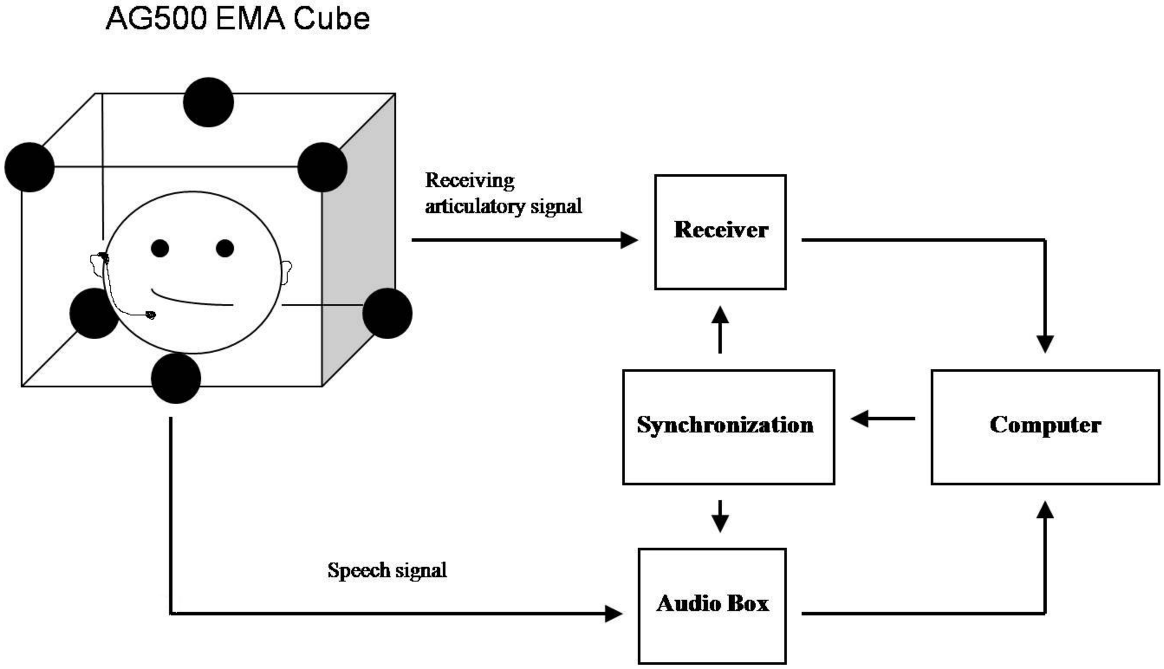


Figure 3.4: Schematic representation of the EMA setup.

Participants were seated on a wooden chair and their heads were positioned within the spherical measuring range of the EMA cube. The filled circles in Figure 3.5 show the approximate locations of the sensors used in this experiment. Sensors were mounted on the tongue tip (TT, 1cm behind apex), the tongue body (TM, 1cm behind the tongue tip sensor) and the tongue back (TB, 1cm behind the tongue center sensor), as well as on the lower incisor (LI), the upper lip (UL), the lower lip (LL), and the right and left corners of the lip in order to track the movements of the tongue, lips and the jaw. Reference sensors were attached to the bridge of the nose and left and right tragi as reference points to normalize head movement.

All stimuli were presented on a laptop computer monitor. Acoustic data was

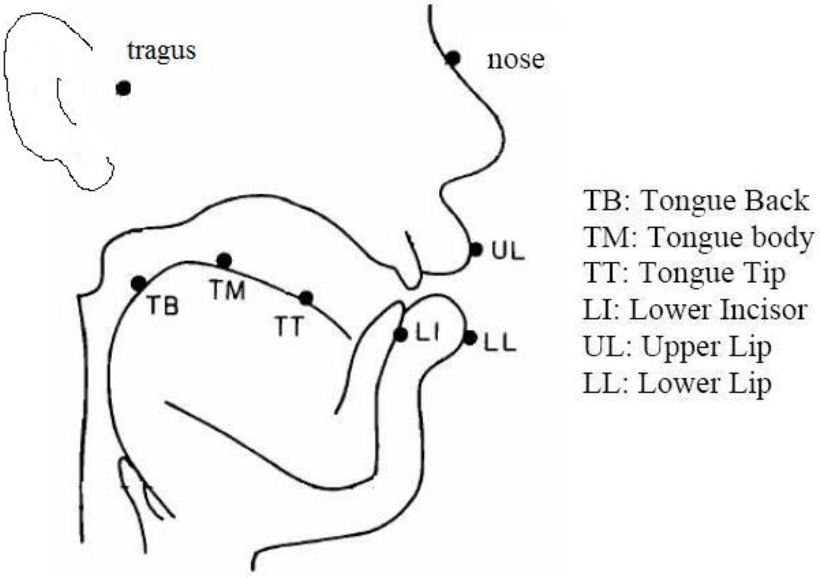


Figure 3.5: The schematic view with three sensors on the tongue, four sensors on lips, and three reference points on nose bridge and tragi.

sampled at 16000 Hz and articulatory data was sampled at 200 Hz. After obtaining the data, the magnetic strength data was converted to positional coordinates. Head movement corrections were carried out so that the movements of the articulators were independent of head movement. After these correction, the measured points corresponded to the three coordinate planes of motion (x; y; z) and to two rotational angles.

Vowel boundaries of all Mandarin and English stimuli were marked automatically using the Penn Phonetics Lab Forced Alinger (P2FA) (Yuan & Liberman, 2008). The author inspected and corrected the boundaries manually to ensure accurate phone segmentation.

3.4.2 Articulatory Analysis

This section presents the articulatory properties of all Mandarin vowels [i, ɨ, ɯ, y, u, e, ɛ, ə, o, ɔ, a, ɑ] from one female native speaker. The purpose of analyzing is to examine vowel categorization from the aspect of articulation, in addition to phonetics and phonology. To do this, the articulation positions and formant frequency at the mid point of the vowel duration were extracted. Auditory judgements of vowel quality were made to verify that the productions were suitable for acoustic analyses. Possible measurement errors with the EMA system and formant tracking with WAVES+ were examined and eliminated from the data analysis. Figure 3.6 (a) shows the mean positions at the points of the tongue body and the tongue tip in the X (anterior and posterior) and Z (superior and inferior) dimensions for vowels [i, y, u, e, ə, o, ɑ]. For the tongue body height, [i] has the highest tongue body position, followed by vowels [y], [e], [u], [ə], [ɑ] and [o]. As for the tongue tip height, [y] has the highest tongue tip position, followed by vowels [i], [e], [ə], [u], [o], and [ɑ]. This result is similar to previous studies (T.-C. Wu & Lin, 1989; Torng, 2000). Interestingly, the vowel [ɑ] has a higher tongue body position than the vowel [o]. This observation is similar to what Torng (2000) found as well. As for the anterior and posterior dimensions, the vowels [i, y, e] have the most anterior tongue body position, followed by the vowels [u]/[a]/[ə] and then [o]. The vowel [y] has the most anterior tongue tip position, followed by the vowels [i]/[e], [u]/[ɑ], and then [ə]/[o].

Figure 3.6 (b) shows the mean positions at the points on the lower tooth, which represent the jaw position, in the X (anterior and posterior) and Z (superior and inferior) dimensions. It is observed that [y] and [u] have the highest jaw

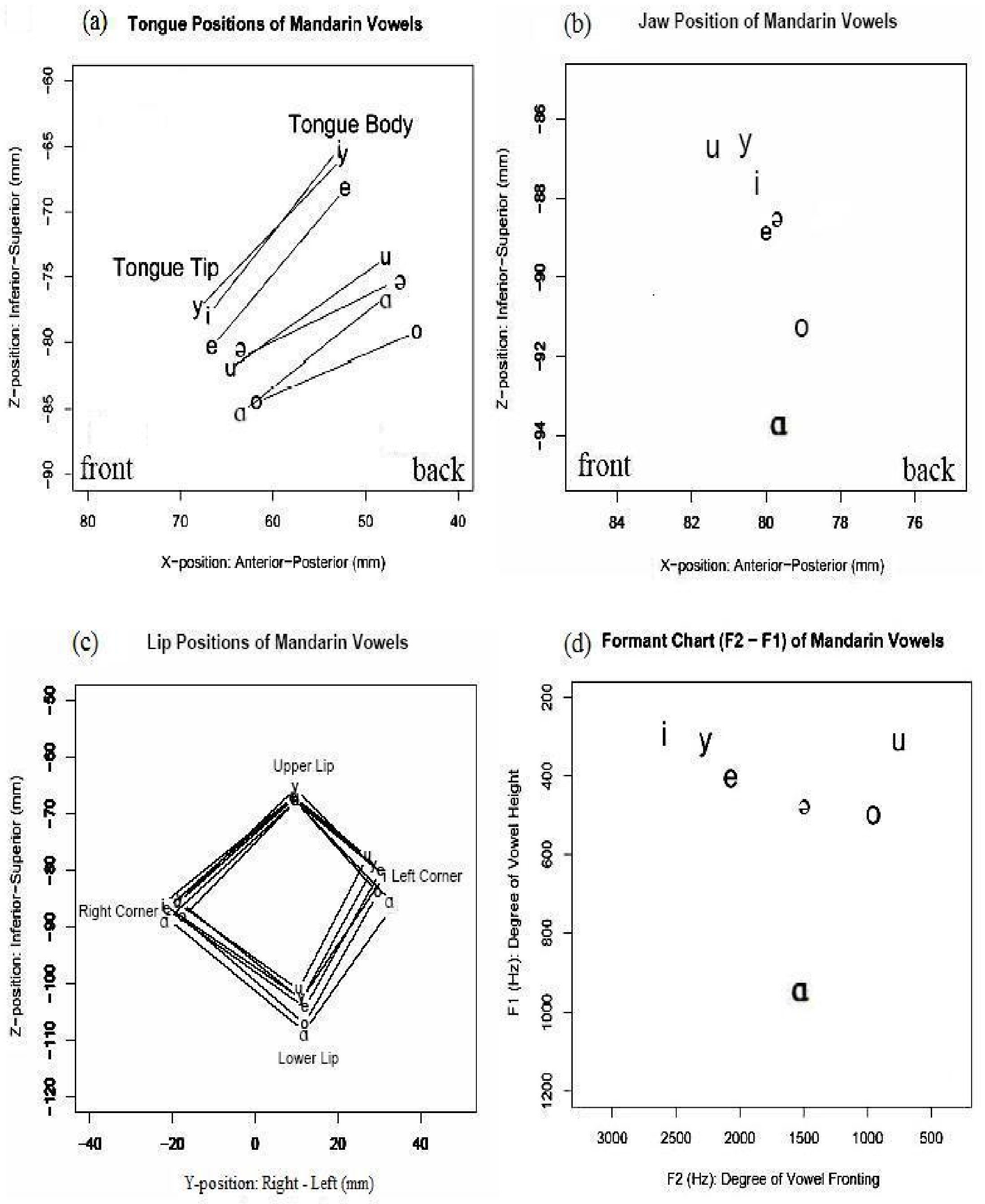


Figure 3.6: (a) shows the tongue position of Mandarin vowels; (b) shows jaw position of Mandarin vowels; (c) shows lip positions of Mandarin vowels; (d) shows a formant chart of Mandarin vowels.

position, followed by vowels [i], [ə]/[e], [o] and [a]. In addition, [y] and [u] have the same jaw position as well as [ə]/[e]. In Wu and Lin (1989), the vowel [i] has the highest jaw position, followed by [y], [u], [o] and [a]. In Torng (2000), the vowels [y], [u], [o] have a high jaw position and the vowel [a] has a low jaw position, while the high vowel [i] has a low jaw position. There is agreement among Wu and Lin (1989), Torng (2000) and the current study that both [y] and [u] have a high jaw position.

Figure 3.6 (c) presents the mean positions of four points, the upper and lower lips and the right and left corners of the lips, in the Y (left and right) and Z (superior and inferior) dimensions. It is noticeable that the vowels [y, u] have a strong protrusion at the lower lip and the right and left corners of the lips, followed by [o]. The vowel [ɑ] has the weakest lip protrusion. The points at the lower lip and the left corner of the lips for vowels [i, e] are slightly more interior than the vowel [a].

The formant analysis of the articulatory data is illustrated in Figure 3.6 (d), which is a vowel formant space plotted by the mean values of the F1 and the F2 at the mid point of each vowel duration. As expected, the vowel [i] has the lowest F1, followed by [y]/[u], [e], [ə]/[o] and [ɑ]. Also, the vowel [i] has the highest F2, followed by vowels [y], [e], [ə]/[ɑ], [o] and [u]. Although the vowels [i] and [y] are difficult to differentiate by F1 and F2, they can be further discriminated by F3 (mean F3 of [i]: 3230 Hz; mean F3 of [y]: 2758 Hz). Table 3.11 shows the first three formant values of these vowels.

Traditionally, the formant values of vowels are usually related to articulatory descriptions of vowels. That is, F1 represents the vowel height as the tongue body height and F2 reflects both the front and back of the vowel as the anterior

Vowel	F1 (Hz)	F2 (Hz)	F3(Hz)
/i/	291	2592	3230
/y/	306	2226	2758
/u/	314	755	2895
/e/	407	2072	2898
/ə/	485	1499	3050
/o/	500	954	3217
/ɑ/	952	1519	2866

Table 3.11: Mean formant values of vowels (Hertz) in Mandarin by one female native speaker

and posterior positions of the tongue and the rounding of the lips. However, the formant values of vowels are determined by the position of the maximum constriction of the vocal tract, which affects the length and the cross-sectional area of the front and back cavities (tubes) in the multi-tube models of vowel production (Stevens & House, 1955; Fant, 1960). In addition, lip protrusion and larynx lowering lengthen the vocal tract, which lowers all formant frequencies (Johnson, 1997, p. 94). Moreover, lip rounding and tongue body raising have the effect of lowering F2 (Perkell, Matthies, Svirsky, & Jordan, 1993).

In the current data, the analysis of the articulatory positions of the tongue and formant frequencies reveals discrepancies between the traditional phonetic descriptions of vowels and the actual tongue positions. Table 3.12 summarizes the vowel quality based on the acoustic data and the tongue positions in the articulatory data. In terms of the acoustic vowel height, the vowel [e] is a mid vowel and the vowel [u] is a high vowel (low F1). However, the vowel [e] has a higher tongue position than the vowel [u]. Similarly, the vowel [ɑ] is a low vowel (high F1) and the vowel [o] is a mid vowel, while the vowel [ɑ] has a higher tongue position than the vowel [o]. In other words, the vowels [e] and [ɑ] have higher tongue body positions than the vowels [u] and [o], respectively. In terms of backness, the vowels [u] and [o] have a more posterior tongue body position than the vowel [e]. The tongue body position in Figure 3.6 (a) does not indicate the place of maximum constriction of the vocal tract. Since we know that vowels [u] and [o] have constrictions both at the lips and at the soft palatal area, the retracted tongue body position and the lip rounding of the vowels [u] and [o] maintain a lower F1 than the vowels [e] and [a], respectively.

	Superior/Inferior (from high to low)	Anterior/Posterior (from front to back)
Acoustics	**F1**: i > y , u > e > ə , o > ɑ	**F2**: i > y> e > ə , ɑ > o > u
Articulation: Tongue body	i > y > e > u > ə > ɑ > o	i , y, e > u , ə , ɑ > o
Articulation: Tongue tip	y > i > e > ə > u > o , ɑ	y > i , e > u > ə , ɑ > o

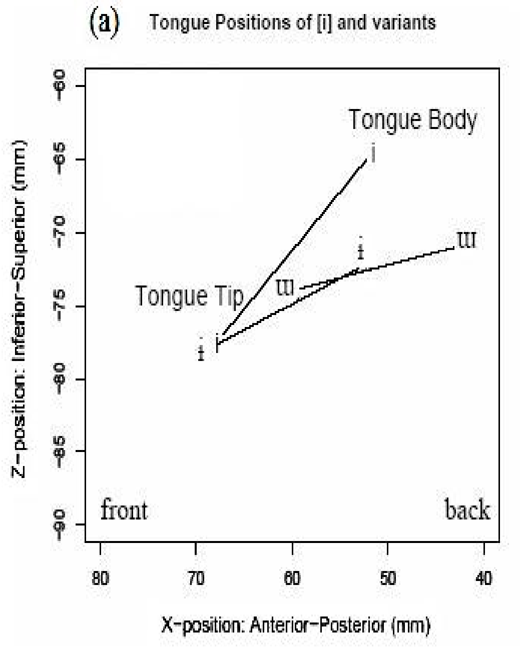
Table 3.12: Comparison between acoustics and articulation

The discrepancies between the articulatory and the acoustic data can be explained. The formant values corresponding to the acoustic vowel quality are influenced by the length of the vocal tract, the constriction point and area function in the vocal tract. Ladefoged (1975) has pointed out that the position of the highest point of the tongue is not a valid indicator of vowel quality. The term ‘vowel height’ refers to an auditory quality that can be specified in the acoustic properties rather than in articulatory positions. In quantal theory (Stevens, 1972, 1989), the relation between the articulatory parameter and acoustic output is not linear, i.e., the acoustic output within some quantal regions is insensitive to the change of the articulatory parameter. In sum, the data here show that the changes in articulation do not necessarily change the acoustic output. The tongue height does not completely correspond to the vowel height in formant values. A question that arose here is: when learners acquire the L2 sound system, do they perceive the articulatory properties of the vowels as PAM (Best, 1995) argues or the phonetic vowel space, as SLM (Flege, 1995b) suggests? When there are discrepancies in vowel categories from different perspectives, such as phonology, acoustics, and articulation, which aspect should be compared to evaluate crosslinguistic segmental similarity?

As discussed in Chapter 3.2, the Mandarin high vowels [i, ɨ, ɯ] are allophones and are grouped as one phonological category under /i/. Phonetically, [ɨ] and [ɯ] are viewed as 'empty vowels' because they extend the articulatory features of the preceding consonants (Ladefoged & Maddieson, 1996). Alternatively, they are treated as one sound of apical vowels following alveolar or post-alveolar sibilants. Moreover, these three vowels are transliterated with the same letter *i*. Due to the debates of the status of [ɨ] and [ɯ], their articulatory positions are examined.

Figure 3.7 (a) shows the mean values of the tongue positions in the X (anterior and posterior) and Z (superior and inferior) dimensions for the vowels [i, ɨ, ɯ]. It is noted that the tongue position of the vowel [ɯ] occurring with the post-alveolar retroflex consonants is further back than that of the vowels [i] and [ɨ]. The vowel [ɨ] has a slightly more front tongue tip position than the vowel [i], whereas the vowel [i] has a higher tongue body position than its variants [ɨ] and [ɯ]. Figure 3.7 (b) and (c) show the formant space plotted by the mean values of F1 as a function of F2 and F2 as a function of F3, respectively. As we can see, the formant values of [ɨ, ɯ] are close to each other, but more distinct from [i]. [ɨ] and [ɯ] have lower F2 and F3 than the vowel [i]. Table 3.13 gives the first three formant values of these vowels.

Although the formant values show that [ɨ] and [ɯ] are similar, the tongue positions of these two vowels are quite different. An X-ray study cited by Cheng



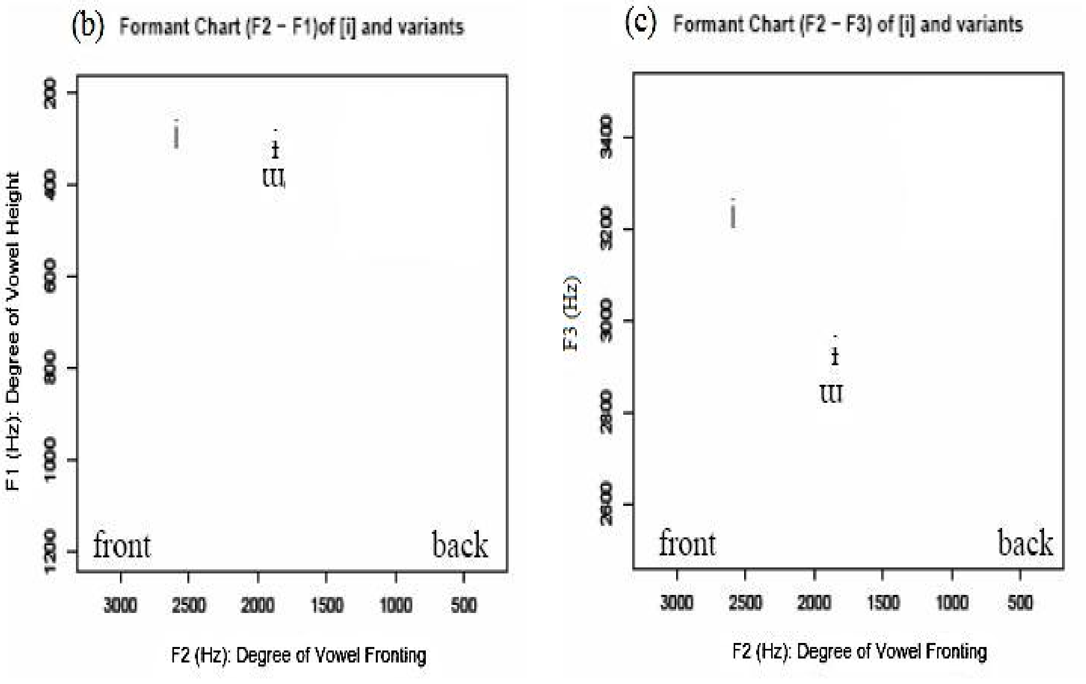


Figure 3.7: (a) shows the tongue position of [i, ɨ, ɯ]; (b) shows a formant chart (F2-F1) of [i, ɨ, ɯ]; (c) shows the the formant chart (F2-F3) of [i, ɨ, ɯ]

Vowel	F1 (Hz)	F2 (Hz)	F3(Hz)
/i/	281	2626	3396
/ɨ/	337	1854	2914
/ɯ/	394	1874	2837

Table 3.13: Mean formant values of [i] and its allophones in Mandarin by one female native speaker

(Cheng, 1973, p. 13) gave the first description of these apical vowels.

> “In X-ray studies, Diàn-fú Zhou and Zōng-jì Wu (1963), comparing these apical vowels with the high front unrounded vowels, find that in the production of the apical vowels the highest point of the tongue is slightly more front and the back of the tongue is slightly higher. Thus these apical vowels have two simultaneous points of articulation, one at the tongue tip and the other at the body of the tongue. Phonetically, the apical articulation may be more distinct, but the articulation of the body of the tongue seems to be more important in terms of phonological patterning.”

If the tongue body position should be the criterion of phonological patterning, [i], [ɨ], [ɯ] would be treated as three distinct vowels. If the last two vowels are not explicitly taught in L2 sound acquisition, learners may lack the concepts of these two vowels or treat them as an “empty” category. Alternatively, they may pronounce [ɨ, ɯ] as [i] due to the transliteration system. All of these issues predict that [ɨ] and [ɯ] are particularly difficult for L2 learners.

The Mandarin low vowel [a], conditioned by alveolar nasal [n] is an allophone of [ɑ]. However, the coarticulation effect has a great influence on vowel quality in Mandarin (Shih, 1995). Figure 3.8 (a) shows the mean values at the positions of the tongue body and the tongue tip for the vowel [ɑ], when followed by [j], [w], [n], and [ŋ] or when not followed by anything. Obviously, the vowel [ɑ] in different contexts has a wide range of formant values and articulatory movements. They are distinct from one another with the exception of the overlap of the monophthong

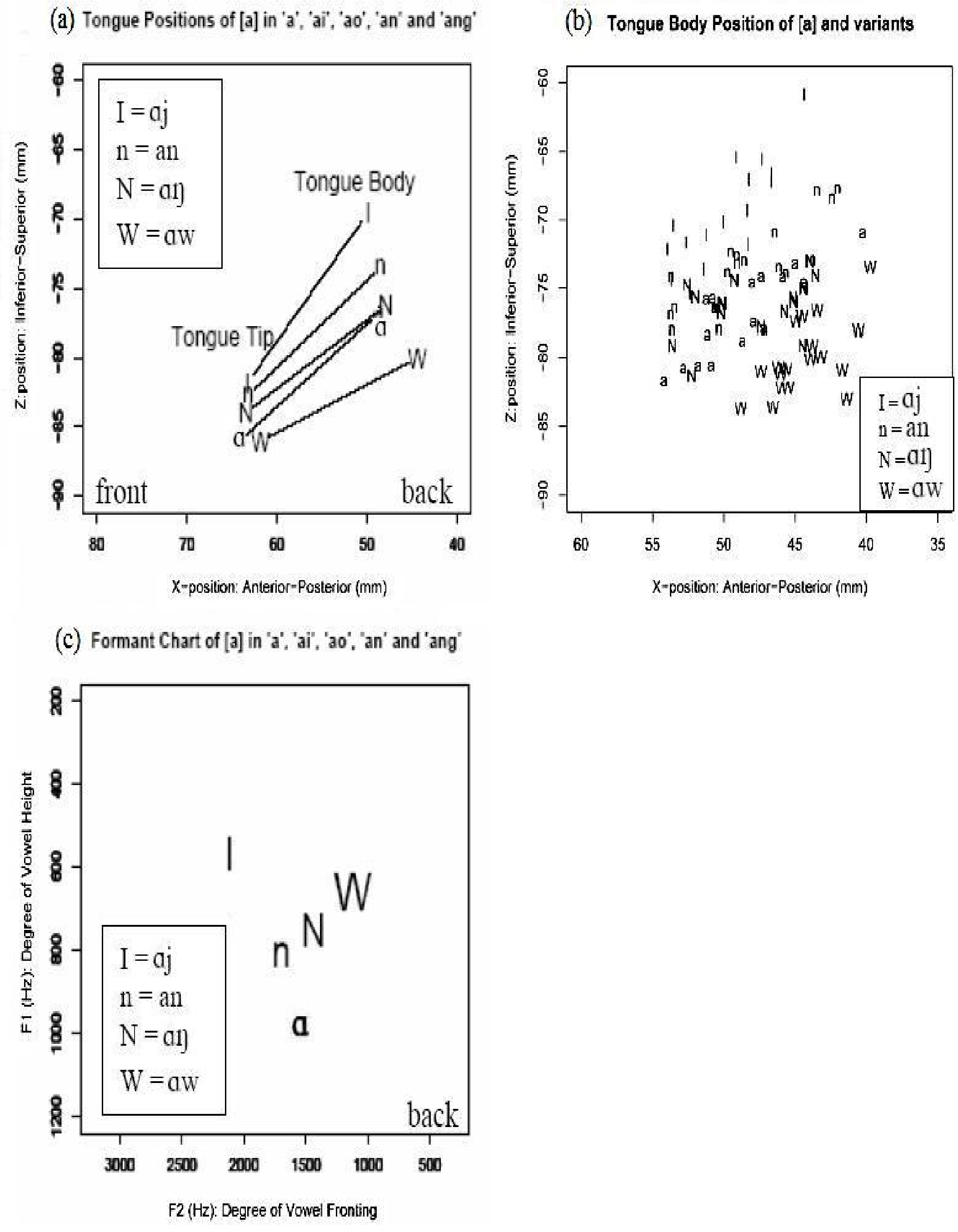


Figure 3.8: (a) shows the tongue position of [ɑ] with its variants; (b) shows the distribution of the tongue body position of [ɑ] and its variants; (c) shows the formant chart of [ɑ] and its variants.

[ɑ] and the vowel [ɑ] in [aŋ] at the tongue body position. The diphthong [ɑj] has the highest tongue positions, followed by the vowel [a] in [an], [ɑŋ], [ɑ] itself, and [ɑw]. The monophthong [ɑ] is slightly more front than other vowels, while the diphthong [ɑw] is a little more back than others. Figure 3.8 (b) shows that the distribution of the tongue body position of the vowel [ɑ] in different contexts is spread out considerably. Figure 3.8 (c) presents the mean formant values. As expected, the vowel [ɑ] in the diphthong [ɑj] has a lower F1 and a higher F2 than the monophthong [ɑ], being raised and fronted by the following glide [j] in the formant plot. By contrast, the vowel [ɑ] in the context of diphthong [ɑw] has both a lower F1 and F2 than the monophthong [ɑ], being raised and backed by the following glide [w]. Both the alveolar and velar nasals have an effect of raising the tongue positions for the vowel [ɑ]. F1 and F2 are consistent with the anticipated tongue position of the following glides. The coarticulation effect observed here supports Shih (1995).

In sum, the data show that the vowel height, in terms of the acoustic measurements, is not equivalent to the tongue height in terms of articulatory positions. Vowel height better reflects the acoustic properties than articulatory movements. The changes in articulation do not necessarily change the acoustic output. The high vowels [i, ɨ, ɯ] can be distinguished by articulation and acoustic properties, while [ɨ] and [ɯ] have similar formant values. The consonantal context has a great influence on vowel [ɑ], which leads to greater variations in tongue positions and formant values. However, learners might not be aware of the change of vowel quality in different contexts and thus use non-native pronunciation. Based on this articulatory study of all Mandarin vowels, the following acoustic investigation of vowels will differentiate the maximum numbers of phonetic distinctions, i.e., five high vowels [i, y, u, ɨ, ɯ], five mid vowels [e, ɛ, ə, o, ɔ] and two low vowels [a, ɑ]. Hence, an exhaustive comparison of vowel similarity between Mandarin and English as well as the comparison of vowel production by Mandarin native speakers and L2 learners in corpus data will be explored.

3.4.3 Acoustic Analysis

For this study, all the monophthongs in Mandarin and English were included for comparison. To avoid dealing with large context effects, limited phonetic contexts across Mandarin and English were selected from the recordings of the EMA data collection for further analysis. In Mandarin, vowels [i, y, u, ɨ, ɯ, e, ɛ, ə, o, ɔ, a, ɑ] in the two contexts of ‘hV’ and ‘bV’ were analyzed because each stimulus in Mandarin only had two repetitions. In English, vowels [i, ɪ, u, ʊ, e, ɛ, æ, ʌ, o, ɔ, ɑ] in the context of ‘hVd’ were chosen. All four repetitions of each word in English were analyzed.

The first three formant frequencies for each vowel were tracked automatically using ESPS Xwaves (*Waves+ Manual version 5.1.*, n.d.). These values were extracted at the midpoint of each vowel duration to avoid formant transitions from the preceding and the following sounds.

Table 3.14 and Table 3.15 list the formant values by gender in Mandarin and English, respectively. A one-way ANOVA comparing the mean formant values by gender in Mandarin and English vowel productions revealed that female speakers have a significantly higher F1 and F2 (Mandarin: F1, F(3, 11) = 2.41, $p < .01$; F2, F(3,11) = 6.52, $p < .01$; English: F1, F(3,10) = 4.90, $p < .01$; F2, F(3,10) = 2.16, $p < .05$). In addition, speaker variability is usually an issue for comparing formant frequencies between groups. Thus, formant values were separated by gender for further analysis.

Post-hoc tests using the Tukey HSD procedure (Winer, 1971) revealed that in the native Mandarin vowel productions of both females and males, the F1 and F2 of [ɨ, ɯ] were significantly different from those of the high vowel [i] ($p < .05$), while the F1 and F2 of [ɨ, ɯ] did not show a significant difference between each other. The low vowels [a] showed significant difference of F2 compared to [ɑ] as well ($p < .001$).

Vowel	F1 (Female)	F1 (Male)	F2(Female)	F2 (Male)
/i/	285 (33)	245 (16)	2632 (477)	2095 (131)
/ɨ/	358 (49)	281 (26)	1905 (118)	1600 (64)
/ɯ/	413 (65)	294 (36)	1874 (81)	1567 (67)
/u/	344 (65)	289 (50)	607 (91)	727 (166)
/y/	324 (40)	271 (26)	2251 (189)	1870 (114)
/e/	423 (56)	317 (43)	2411 (539)	2006 (151)
/ɛ/	565 (64)	386 (25)	2107 (365)	1782 (112)
/ə/	613 (57)	421 (37)	1359 (96)	1263 (56)
/o/	434 (73)	353 (33)	782 (66)	716 (125)
/ɔ/	566 (88)	397 (19)	990 (131)	882 (153)
/a/	892 (258)	691 (134)	1753 (110)	1499 (98)
/ɑ/	888 (300)	782 (103)	1476 (106)	1195 (48)

Table 3.14: Mean formant values of vowels (Hertz) in Mandarin by females and males

An ANOVA test of formant values (F1 or F2) to formant values in English, examining the Hertz difference values, yielded a significant interaction (female: $F(4,10) = 6.97$, $p < .001$; male: $F(4,10) = 21.07$, $p < .001$). *Post-hoc* tests using the Tukey HSD procedure confirmed that F2 values were significantly greater for English vowels [i, u, e, o, ɔ] than the Mandarin counterparts produced by male speakers. In female productions, F2 values were significantly greater for English [u, ɔ] than that of Mandarin vowels. The English vowel [ɔ] significantly differs from that in Mandarin in both F1 and F2 for female and male productions. Table 3.16

Vowel	F1 (Female)	F1 (Male)	F2(Female)	F2 (Male)
/i/	314 (53)	252 (38)	2756 (183)	2405 (102)
/ɪ/	556 (118)	440 (43)	2151 (279)	1969 (115)
/u/	351 (48)	303 (39)	1216 (213)	1019 (140)
/ʊ/	568 (84)	466 (66)	1601 (96)	1448 (108)
/e/	407 (21)	356 (53)	2652 (154)	2310 (150)
/ɛ/	693 (117)	595 (40)	2087 (116)	1816 (69)
/æ/	940 (133)	785 (36)	1898 (123)	1642 (64)
/ʌ/	680 (142)	557 (57)	1815 (165)	1479 (68)
/o/	425 (60)	412 (64)	1069 (161)	1055 (454)
/ɔ/	866 (125)	717 (130)	1416 (254)	1266 (403)
/ɑ/	929 (132)	650 (173)	1508 (290)	1149 (147)

Table 3.15: Mean formant values of vowels (Hertz) in English by females and males

lists the output of the ANOVA analysis between Mandarin and English vowels.

Figure 3.9 shows the scatter plots of the distribution of vowels in Mandarin and English, by gender. The X axis indicates F2 (Hz), representing the backness of vowel quality and the Y axis indicates F1 (Hz), representing vowel height. Languages are color-coded, blue for Mandarin and red for English. Figure 3.10 illustrates the mean formant values of vowels. Note that the major difference is the greater F2 values of English vowels, meaning that Mandarin vowels are further backed than English vowels in terms of acoustics. Observations combining the statistical results and distribution of vowels can be summarized as listed below:

- Mandarin [i], [e] and [o] are statistically identical to their English counterpart in female production. They show a difference in the male productions of F2.
- Mandarin vowels [ɨ, ɯ] are statistically different from English vowels [i] in terms of F2 in both female and male productions.
- Mandarin vowels [ɨ, ɯ] are statistically different from English vowels [ɪ] in both F1 and F2 in both female and male productions.
- Mandarin [u] is statistically further back than the English counterpart in both female and male productions.

- Mandarin and English [ɛ] have no significant difference of F1 and F2 in female productions, but are significantly different in F1 in male productions. Further, the formant charts show that the distribution of [ɛ] in these two languages are different.
- Mandarin [ɔ] is significantly different from the English counterpart in the F1 and F2 of both female and male productions.
- Mandarin [a] and [ɑ] have no significant difference from the English [ɑ] in both F1 and F2 in female productions.
- In male productions, Mandarin [ɑ] is statistically different from the English [ɑ] in F1, while Mandarin [a] is significant different from the English [ɑ] in F2.

The observations in this study agree with the findings of Thomson et al. (2009), which reported that the most similar vowel in Mandarin and English is [o], and second most similar vowels are [ɑ, i, e]. The most different vowels are English [ɪ, ɛ, u], which are not similar to any Mandarin vowel categories.

	Female		Male	
Mandarin - English	F1 (Female)	F2 (Female)	F1(Male)	F2 (Male)
/i/ - /i/	-	-	-	***
/ɨ/ - /i/	-	***	-	***
/ɯ/ - /i/	-	***	-	***
/ɨ/ - ɪ/	***	-	***	***
/ɯ/ - ɪ/	*	*	***	***
/u/ - /u/	-	***	-	***
/y/ - /u/	-	***	-	***
/e/ - /e/	-	-	-	***
/e/ - /ɛ/	***	*	***	-
/e/ - /æ/	***	***	***	***
/ɛ/ - /ɛ/	-	-	***	-
/ɛ/ - /e/	-	***	-	***
/ɛ/ - /æ/	***	-	***	-
/ə/ - /ʌ/	-	***	***	-
/o/ - /o/	-	-	-	***
/o/ - /ɔ/	***	***	***	***
/ɔ/ - /o/	-	-	-	-
/ɔ/ - /ɔ/	***	***	***	***
/ɑ/ - /ɑ/	-	-	***	-
/a/ - /ɑ/	-	-	-	***
/ɑ/ - /æ/	-	***	-	***
/a/ - /æ/	-	-	-	-

Table 3.16: Comparison between Mandarin vowels and English vowels, $^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$

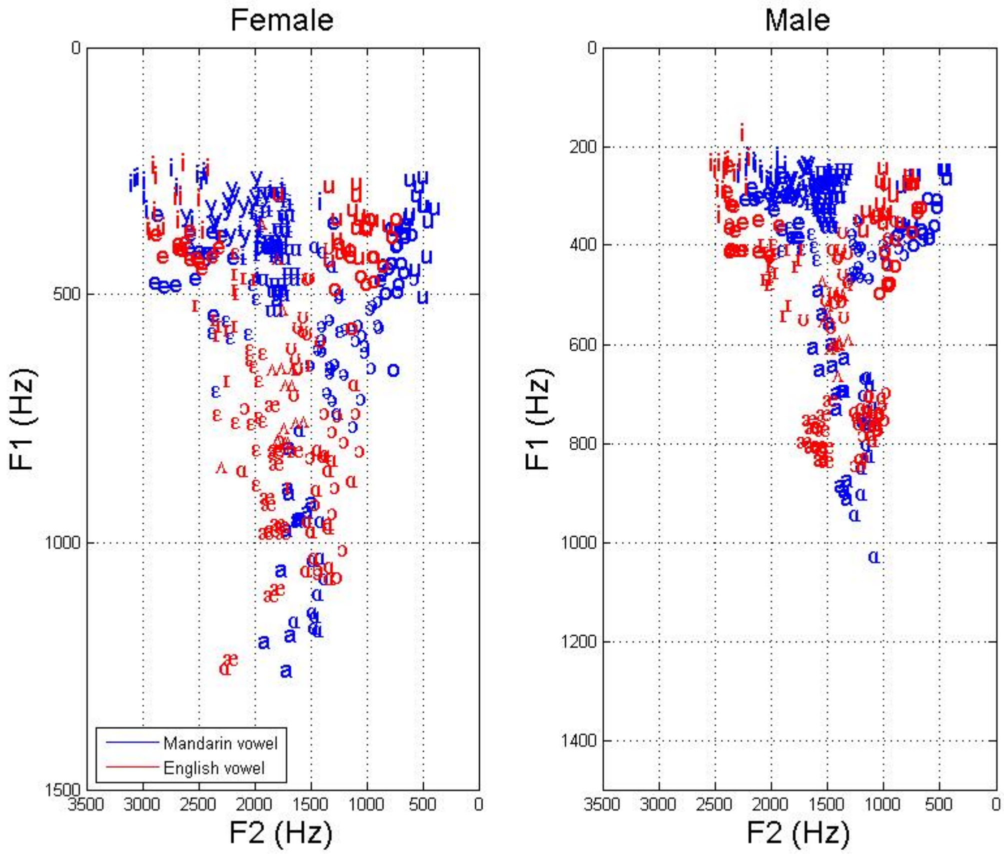


Figure 3.9: Vowel distribution in Mandarin and English by native speakers.

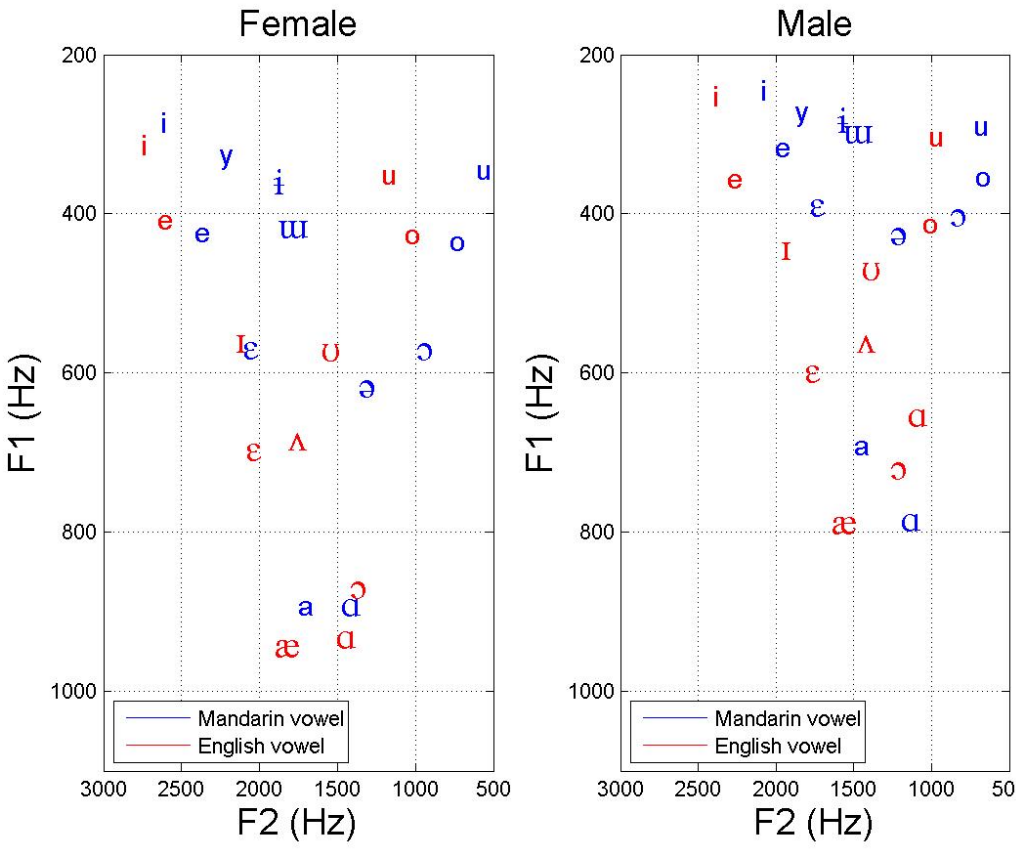


Figure 3.10: Vowel Space of mean formant values in Mandarin and English by native speakers.

3.4.4 Discussion

The Mandarin vowel inventory is a controversial topic. Phonologically, many vowels are predictable from context. i.e., they occur in complementary distribution. Without minimal pairs, it is not a straightforward task to provide evidence of vowel categories in either the phonological input or the phonological output. For instance, whether the apical vowels [ɨ, ɯ] exist in surface forms due to their limited distribution is debatable. With little agreement in the literature, we do not know the maximum number of Mandarin vowels, though the number of possible rhymes is generally agreed on. We also do not know whether native Mandarin speakers treat some of these vowels as the same, and if so, which ones should be combined. However, this interesting question is beyond the scope of this thesis.

In order to fully examine the vowel similarities between Mandarin and English and predict L2 learners' behavior, I took all phonetic categories of Mandarin vowels into account. The Mandarin vowels [i, y, u, ɨ, ɯ, e, ɛ, ə, o, ɔ, a, ɑ] were compare to the English vowels, [i, ɪ, u, ʊ, e, ɛ, æ, ʌ, o, ɔ, ɑ]. In native Mandarin vowel productions, it was found that [i, ɨ, ɯ] are articulated different through tongue positions. Phonetically, [ɨ, ɯ] were significantly different from [i] in the F2 domain, but [ɨ] and [ɯ] themselves have similar formant values. Also, the pair [ɑ] and [a] demonstrated different articulatory properties due to the coarticulation effect of preceding glides or following coda nasals. Statistically, they significantly differ from the F2 of native production.

The acoustic comparison between Mandarin and English vowels was investigated and it was demonstrated that the primary discrepancy emerges from the further backed vowel quality in Mandarin (lower F2). Detailed vowel-to-vowel comparison was shown in the previous section. According to the major SLA theories, assessing crosslinguistic segmental similarities is important for predicting L2 sound acquisition. According to SLM's definition of similarities and differences in sounds, Mandarin vowels can be classified into four categories: different vowels, new vowels, identical vowels, and similar vowels. The different vowel is [ɔ] because it exists in both languages, but shows significant differences in both F1 and F2 in female and male productions. The new vowels are [y, ɨ, ɯ] due to the fact that they do not exist in English. The identical and similar vowels pattern differently in female and male production. In female productions, the identical vowels are the vowels with no statistical different in F1 and F2, while the similar vowels show an F2 difference. In male productions, there is no distinction between identical and similar vowels because all of them show at least either F1 or F2 differences.

- The different vowel is: [O].
- The new vowels are: [y, ɨ, ɯ].
- The identical vowels between Mandarin and English are: [i, e, ɛ, o, a, ɑ].
- The similar, but not identical vowels are: [u, ə].

The SLM predicts that similar vowels are difficult to learn, while it is easier to establish new categories for different vowels. Hence, learning Mandarin [y, ɨ, ɯ, ɔ] should be relatively easier than learning Mandarin [u, ə]. If Mandarin [i, e, ɛ, o, a, ɑ] are treated as identical to the English counterparts, L2 learners should not have any problems in acquiring these sounds.

3.5 Summary

In this chapter, I reviewed how the vowels in Mandarin have been analyzed from different perspectives, including phonological categories, phonetic measurements and articulatory investigations in the literature. Unlike Mandarin tones, it is not clear how many Mandarin vowels exist in the underlying representation and the surface forms. According to various theoretical approaches, the number of Mandarin vowels ranges from two to six in the underlying representation and nine to fifteen in the surface forms (Cheng, 1973; Y.-H. Lin, 1989; Wang, 1993; Duanmu, 2000). Although this issue is still in dispute, I adopted the maximum number of phonetic monophthongs in Mandarin for further analysis. In addition, the learning problem caused by the transliteration system, Pinyin, as well as the spelling confusion between Mandarin and English were discussed. The methods of assessing crosslinguistic segmental similarities, including perceptual judgements, spectral comparison of the extent of the overlap distribution and a statistical pattern recognition model were presented. By using EMA, the articulatory properties of Mandarin vowels were investigated. Also, phonetic measurements provided the acoustic values to evaluate the similarity and dissimilarity between Mandarin and English vowels. Based on SLM's definition, Mandarin vowels were grouped into four types to predict the learning behaviour. In chapter 4, the vowel productions of Mandarin natives, Chinese heritage speakers and English speaking learners of Mandarin in the corpus data will be used to test the SLM's prediction.

Chapter 4

Study of Fluency and Foreign Accent

In this chapter, two corpora are used to study second language fluency and foreign accent. First, I introduce the corpora and the data management. Second, the experiment design of sampling selection and perceptual ratings are explained. The analyses, including rating results, principal component analysis, factor analysis, acoustic analysis of speech properties and vowels analysis are reported.

4.1 Introduction

In this chapter, I am going to use two large corpora to investigate questions about the production and perception of second language fluency and foreign accent. The first corpus is the Spontaneous Chinese Learner Speech Corpus, which was a longitudinal study of classroom speech production recorded from 2004 to 2009 at the University of Illinois, Urbana-Champaign (UIUC) (Shih, 2006; Shih & Wu, 2011). The corpus includes native speakers (instructors) and learners (both heritage and non-heritage students). Their speech shows variations of fluency levels and degree of accents. The other corpus is a picture telling corpus, where subjects performed tasks including clock telling, simple picture description and complex picture description (Bock, Irwin, Davidson, & Levelt, 2003; Griffin &

Bock, 2000; Papajohn, 1998). The advantage of running a parallel analysis on two different corpora is the ability to see whether findings can be replicated. Replication is important in science because it helps to assure the results are not skewed or obtained by chance. Doing parallel analysis on different data allows us to validate the perceptual rating results.

The speech data in this study includes spontaneous and prepared speech which have been produced in a natural setting and presents phenomenon that may not be observed in experiments of reading word lists. Fluency and foreign accent may exist in spontaneous conversation, while the properties contributing to fluency and foreign accent might not be prominent in read speech. Random sample selection from large corpora is used to avoid sampling bias and the results are representative of the corpora and generalizable to the population that the corpora represent.

Due to the development in computational power, networks and computer storage, analyzing large amounts of spontaneous speech has recently become a possible task. What we report will be a novel attempt to observe the relationship between acoustic attributes related to fluency and foreign accent and human-assigned perceptual ratings. Further, predictions, based on vowel similarities, defined by the SLM can be tested by examining the vowel productions of different speaker groups in the corpus data. The rating metrics investigated in this study include questions of fluency, nativeness, accentedness, disfluencies, pronunciation, vocabulary, grammar and comprehensibility. Fluency reflects the degree of speech smoothness. Nativeness is viewed from the perceptual side to see if a speaker sounds like a native speaker. Accentedness is defined as the perceived distance of speech production between the speaker and listener. Disfluencies are related to the detection of hesitation, repairs, pauses and silence in speech. Pronunciation is usually used in testing or in classroom settings to evaluate the level of a learner's oral proficiency. Grammar is used to gain a sense of the perceived accuracy of sentence structure in speech. Comprehensibilty is a measure of the intelligibility of the delivered message.

Applying these questions to a large, randomly selected speech samples collected in natural environment, we hope to advance our understanding of second language fluency and foreign accent. The SLA theories can be tested and the contribution of vowel pronunciation to foreign accent can be examined in detail.

4.2 Spontaneous Chinese Learner Speech Corpus

The corpus consists of 185 hours of audio and video recordings from the third-year and fourth-year Chinese language classes, which was recorded in a Chinese speech training class on a weekly basis from Fall 2004 through Spring 2009 at UIUC (Shih, 2006; Shih & Wu, 2011). Speaker background in this corpus varies (e.g. teachers, heritage learners, non-native L2 learners with different L1 backgrounds). Hence, this database is a prolific resource with speech samples representing various spectra of fluency and foreign accent. Before the corpus data can be used for various research topics, such as perceptual ratings, language assessment, acoustic analysis, speech recognition etc., it requires many layers of labor-intensive work. Different research topics all require an appropriate unit for speech sample selection and data processing. Thus, the first line of work is to mark speaker turns. This step also provides speaker codes and the precise time boundaries demarcating the hour-long recordings into speech turns. Based on the turn-markings, each snippet was displayed on a webpage to obtain a turn-sychronized transcription. A subset of the data was selected for acoustic analysis and perceptual judgements of fluency and foreign accent. The general properties of the corpora and each step of the data management is explained in the following sections.

4.2.1 Speakers

The speakers in the corpus includes 11 Chinese teachers (9 females and 2 males), 86 Chinese heritage learners (28 females and 58 males), 11 Korean learners of Chinese (7 females and 4 males) and 23 English learners of Chinese (8 females and 15 males). The eleven Chinese instructors are Mandarin native speakers; among them, five are from Taiwan and six from Mainland China.

The heritage learners were students whose native language is Chinese, but who received education in English and grew up in the United States. Some were born in the U.S. and some arrived in the U.S. at a young age . Most of the Chinese heritage learners were from Mainland China and some were from Taiwan. Recently, the language learning development of heritage speakers has attracted the attention of many researchers in SLA (Au, Knightly, Jun, & Oh, 2002; Montrul, 2006, 2008; Polinsky, 2006). Heritage speakers are adult early bilinguals of minority languages. They might be the children of first generation immigrants or might have lived in an L2 country at some point during childhood. Under these conditions, the heritage language might not be completely acquired due to the fact that children of first generation immigrants have a strong desire to fit into the new society. These speakers also speak the native language on a limited basis and in a restricted environment. Therefore, the heritage language used at home might gradually be dominated by the majority language of the new society. The competence and performance of heritage speakers varies to different degrees because of incomplete L1 acquisition. Generally speaking, they have good speaking and listening abilities, with native-like pronunciation and fluency. An interesting question has been raised as to whether heritage speakers perform more like native speakers or adult L2 learners. Although L1 acquisition might not be completely acquired in heritage speakers' childhoods, some of them might go back to L2 classes to improve or maintain their L1.

The non-native L2 learners are students whose native language is English or Korean. The English learners of Chinese are learners who had no prior background in Chinese before they attended the college-level Chinese classes. Different from English learners of Chinese, most of the Korean learners had prior background in Chinese during their high-school education.

4.2.2 Task

Students in the Chinese classes received speech training in two paradigms, namely, "Variety Show" and "Debate" (Shih, 2006). Each of the paradigms was designed to fit in a 50-minute class. In the *Variety Show* format, there are 4 main sessions: opening, talk show, formal speech and comments. Learners are asked to play roles, such as to be the chair for the whole show, to be the talk show host, or to be the speech makers. The chair opens each session in the show with an introduction; the talk show host prepares several topics and selects students from the audience to step up in front of the stage and answer questions. The speech makers give prepared, formal, 4 to 6 minute speeches. The *Variety Show* format incorporates a few frequently encountered social interactions, such as giving an opening/closing remarks, introducing guest speakers, and giving formal speeches. Through weekly practices, learners had multiple chances to play each role and to observe many performances by their classmates and instructors.

In the *Debate* format, students are divided into two sides, a proposition side and an opposition side. A specific topic is given in advance. Some of the learners prepare a formal speech to express their positions on the given topic; some prepare questions to ask the opposing side; and some have to answer questions on the spot. The *Debate* format trains learners to argue and speak clearly, logically and convincingly under time pressure.

Based on different formats, there are two speech styles, namely: (1) spontaneous speech in which students speak without advanced preparation, i.e., some questions and all answers in the Variety Shows and Debates; and (2) prepared speech, i.e., speeches made by the chair or host, the formal speeches prepared by students in Variety Shows and the statements students made in Debates. If students prepared their speeches beforehand, most of them read their speech and held drafts in their hands. Thus, prepared speech can be recognized in the video clips. Overall, the speech style of Debate is more formal than that of the Variety Show.

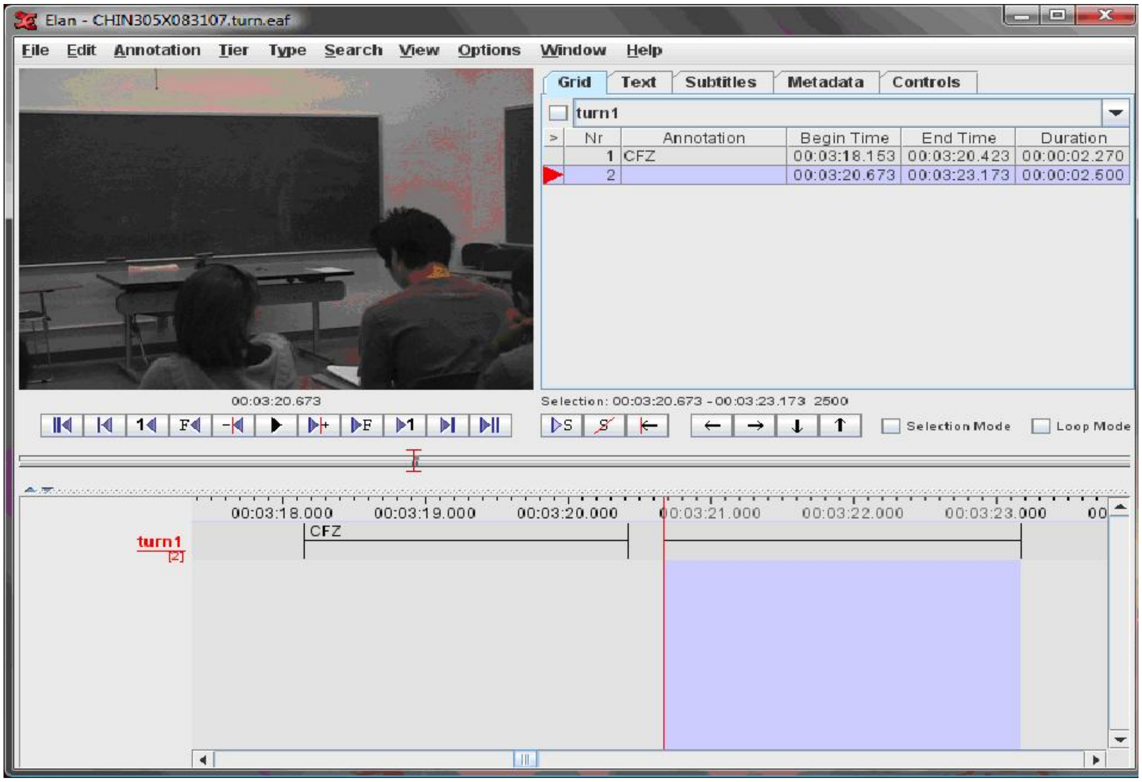
4.2.3 Speaker Turn

Speaker turn-marking is a labor-intensive and time-consuming step in the procedure of database management. The goal of turn-marking is to provide a proper unit for speech sample selection and time-synchronized transcription. Speaker turn-markings were annotated with a precise start time and end time for each turn and identified with speaker codes. There are several reasons why turn-marking is necessary for future research projects.

- An individual speaker's turn, as a unit, facilitates speech sampling for individual speakers. The annotation of speaker codes enables researchers to create pools of speech samples from the same speakers and do sample selection. If speech is overlapped by multiple speakers in a speech snippet, for instance, it will be difficult for raters to evaluate speech. It also increases the difficulty in acoustic analysis.

- Discourse coherent speech is preferred in perceptual human-rated experiments. Speech samples starting from a random point in the speech might bias fluency or accentedness scores. With speaker turn-marking, we are able to select speech samples from the beginning of turns.

- Long speech files increase the chance of misalignment between speech and text when using automatic speech recognition technology (ASR).



>	Nr	Annotation	Begin Time	End Time	Duration
	1	CFZ	00:03:18.153	00:03:20.423	00:00:02.270
	2		00:03:20.673	00:03:23.173	00:00:02.500

Selection: 00:03:20.673 - 00:03:23.173 2500

Figure 4.1: An example of ELAN annotation

In this procedure, speaker turns were marked and annotated by trained research assistants using the video editing software ELAN (Hellwig, n.d.). An example of ELAN annotation is given in Figure 4.1. The annotator uploaded classroom video and audio into ELAN, dropped cursors to indicate the precise time of turn boundaries and entered the speaker code (e.g. 'CFZ' in Figure 4.1). The 50-minute recordings were thus demarcated with speaker codes and time stamps indicating the precise turn boundaries of their speech. The information enables synchronization of speech and transcription and facilitates sampling speech for perceptual ratings, acoustic measures, and ASR training.

Figure 4.2 schematizes the database management in a systematic way. This study started by collecting data and went through the procedure to complete the speaker turn-markings and transcriptions for the entire database in order to establish a sampling frame.

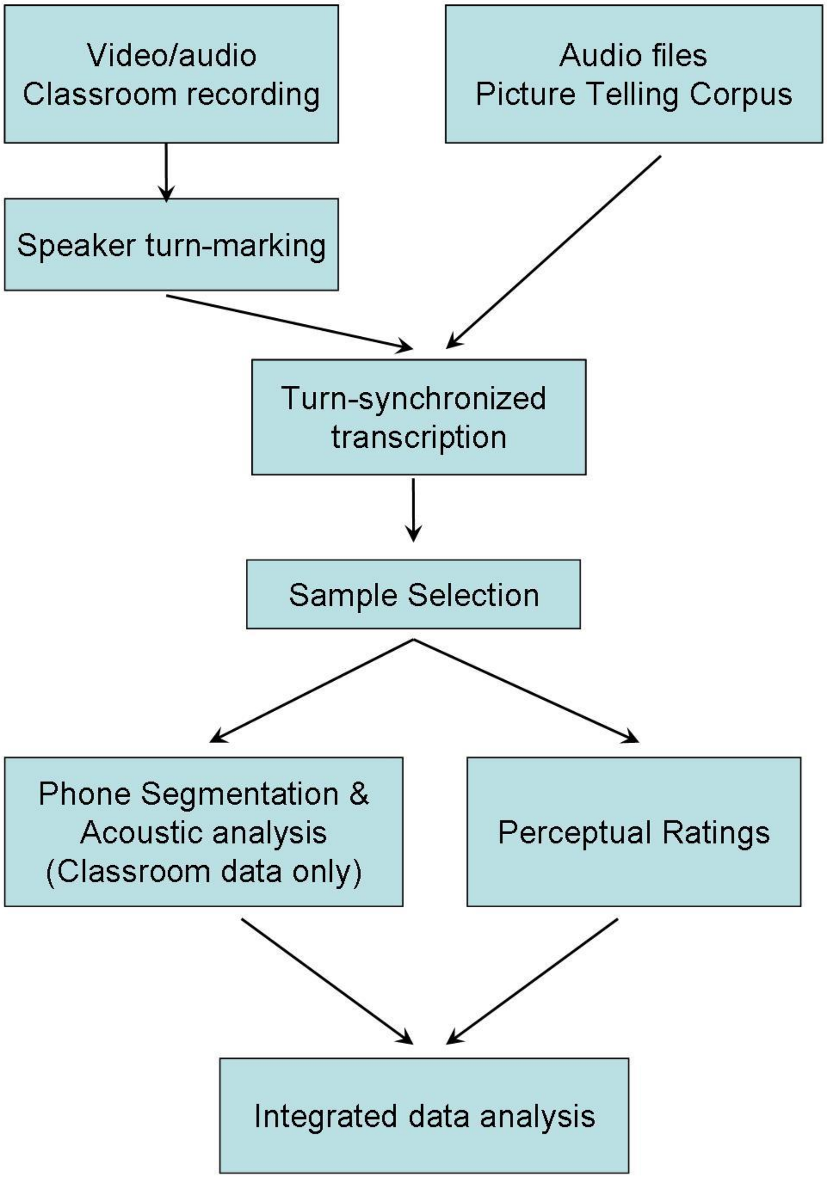


Figure 4.2: Flowchart of the database management.

4.3 Picture Telling Corpus

The picture telling tasks were conducted in the department of Psychology at UIUC in order to study the relationship between L1 fluency and L2 fluency (Bock et al., 2010). Students in the Chinese classes were recruited to participate in this experiment. Subjects were asked to produce speech in three different task types in Mandarin and English. Details of the picture telling experiment are described in the following sections.

4.3.1 Speakers

Nineteen heritage and four English learners participated in the picture telling task. They participated in the classroom recordings as well.

4.3.2 Task

Three tasks namely, clock telling, simple picture description, and complex picture description were developed by Bock and her colleagues (Bock et al., 2003; Griffin & Bock, 2000). In the clock-telling task, subjects were asked to produce either relative time expression (“ten past five”) or absolute time expression (“five ten”) as rapidly as possible after viewing either analog or digital clock faces. In the task of simple picture description, subjects examined a single picture and were asked to describe the content in the picture as much as they could. In the task of complex picture description, subjects were presented with a series of cartoons of six frames one by one and were asked to narrat the story (Papajohn, 1998). Subjects in these three tasks were asked to produce spontaneous speech in English, their L1, and in Mandarin, their L2. Sample pictures of the tasks, clock telling, simple picture description and complex picture description are shown in Figure 4.3.

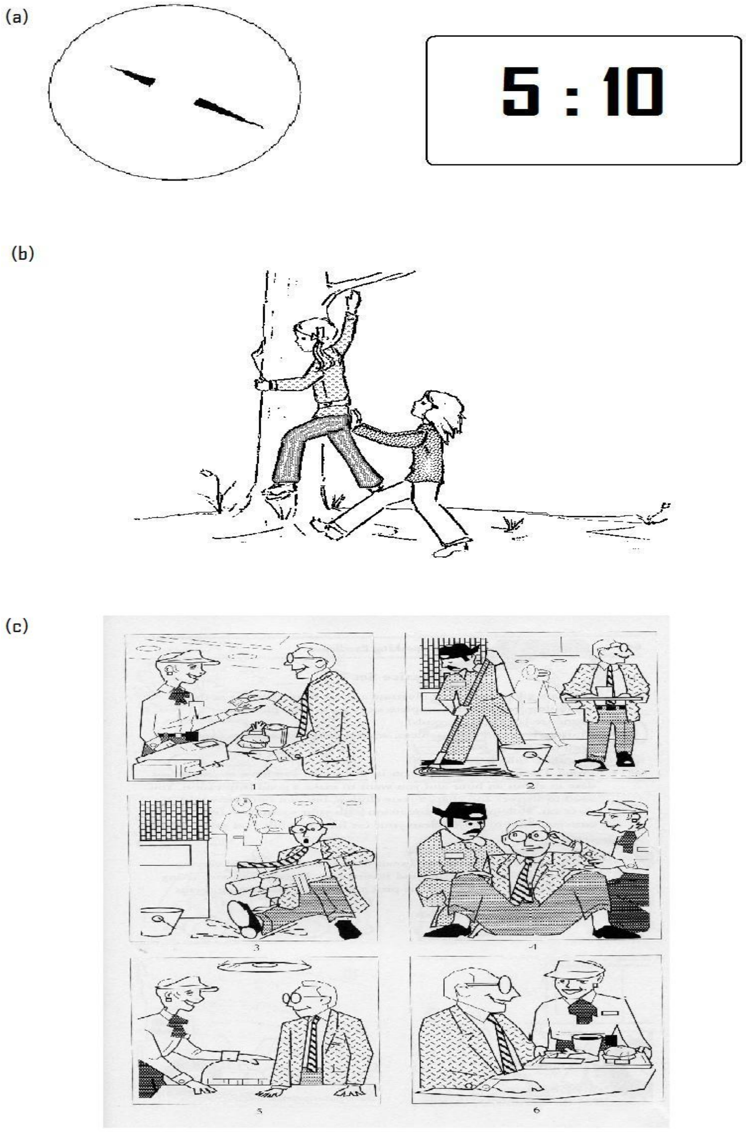


Figure 4.3: Sample picture of the tasks (a) clock telling; (b) simple picture description; (c) complex picture description: picture source (Papajohn, 1998).

4.4 Transcription

The transcription was obtained through a transcription website, where each speaker turn was presented individually with a link to the audio/video files. A text area is provided for verbatim transcribing. The transcribers can make edits and revisions and all versions of the revisions were saved. Speech data from both the classroom recordings and the picture telling corpus were transcribed. The transcription was done using traditional Chinese characters by trained transcribers. Speech recordings produced in English in the picture telling tasks were transcribed in English. A snapshot of the transcription website is given in Figure 4.4.

翻譯標記 片段: CHIN306X031706/segment001

大家好,我是<:>,我是今天的主持人,<嗯>禮拜五<啊>終於到現在已經週末了,那我,我就覺得外面的人<COUGH>很俗,外面天氣那麼好,為什麼要怎樣不在,那我就問我的室友,<啊>他告訴我個很重要的事情，可能不知道,但是,下禮拜,不用上課，<LAUGH>所以沒有人在,其實整個禮拜都不用上課,為什麼,誰知道,<呃>沒有人知道啊,<啊>都完蛋了<LAUGH>下禮拜放假,我知道你現在一定急著要出門,不想在這裡，所以<呃>今天的,短一點就好了,每個人都要回家,或者去玩,所以,今天的主提人是<:>

Figure 4.4: A snapshot of the transcription website

Specific linguistic and non-linguistic phenomena, such as disfluencies, speech errors and laughter were labelled with a pair of angle brackets <>. Below are the transcription guidelines.

- Non-linguistic events, such as laughter, claps, coughes, and other loud noises are transcribed as <LAUGH> ,<CLAP> ,<COUGH> , and <NOISE>, respectively.

- Filled pauses are annotated with corresponding Chinese characters, such as < 嗯 >, < 哦 >, and < 呃 >.
- Unclear speech is labelled with <SKIP>.
- English in speech is transcribed in English enclosed in a pair of angle brackets. If it is not understandable, then it is marked as <ENG> .
- If speech overlaps significantly, the speech is not transcribed but is instead tagged as <OVERLAP>.
- Speech errors are annotated with the expression of <erroneous syllable/intended syllable>. The erroneous syllables or actual spoken sounds are transcribed in Zhuyin, followed by the intended and correct Chinese character, separated by a slash and all enclosed in a pair of angle brackets, for example, < ㄉㄧㄢ 3/等 > 一下. This is a case where the speaker intends to say 等一下 'deng3 yi1 xia4' 'wait a minute', but she said ㄉㄧㄢ 3 一下 'dian3 yi1 xia4'. The numeral indicates the Chinese tone. Learners' systematic pronunciation errors are not annotated.

An example of the transcription is provided as below.

< 嗯 >, 大家好。大家好。非常高興大家今天 < ㄋ > 來參加這一個活動。我先簡單的自我介紹一下。我叫做 <WZH>, 我是四年級中文課的 <TA> 。那今天我們要舉行一場辯論的活動。辯論的題目是, 中國傳統應該保持。坐在這邊的五位是我們正方的, 參加同學, 然後在這邊左手邊的是我們, 五位, 還有一位還沒來, 我們, 的, 反方的同學。那, 在這裏我先簡單的講一下我們 < ㄉㄧㄢ 3/等 > 一下辯論進行的方式。<CLAP>.

<uh> Hello, everyone. Hello, everyone. I am very glad that all of you came to join this activity. Let me briefly introduce myself. My name is <WZH>, I am a fourth-year Chinese <TA>. Today, we are going to have a debate. The topic of the debate is "Chinese traditions should be preserved". Sitting here are the five students of the proposition side, while the other five sitting at the left-hand side are the students of the opposition side. There is another who has not come yet. Well, here I am going to briefly explain the procedure of the debate. <CLAP>.

- <WZH> is the annotation of speaker code.

- <TA> is the annotation of English words when the speaker said 'TA'.
- < ㄉㄧㄢ 3/等 > is the annotation of speech error.
- <CLAP> is the clapping after this segment of speech.

4.5 Method

This study uses a stratified sample design to select speech data of the same speakers from different times so that each speaker had multiple samples adding up to at least one-minute of speech. Using more speech samples from the same speakers is more representative than relying on one sample to observe the speaker's performance. A fairly large number (43) of naïve raters was recruited to evaluate speech samples with a binary rating scale of fluency and a 4-point rating scale of foreign accent and other criteria.

4.5.1 Sampling Design

Good sampling design is an important aspect of research which can lead to reliable statistical inference and predictions. In this study, fluency is determined by how speakers handle connected sentences. Thus, speech samples should be long enough to allow raters to evaluate fluency and foreign accents. In addition, each speech sample should include the inclusion of multiple sentences if possible, rather than being restricted to sentence fragments or single sentences. However, there is no universal agreed-upon length of speech samples for perceptual ratings. While the ACTFL oral proficiency test uses interviews that may be more than 30 minutes, shorter samples have been used successfully in evaluation tasks. A study by Ambady and Rosenthal (1993) demonstrated that student's ratings of instructor's nonverbal behaviors based on 30 seconds of silent video clips composed of three 10 seconds clips from the same teacher, or even thinner slices of 6 seconds and 15 seconds, successfully predicted end-of-semester teaching evaluation. This finding suggested that impressions can be formed extremely quickly. Derwing (2006) used 20-second speech samples for evaluating fluency and foreign accent and observed that 20 seconds was sufficient for raters to make reliable judgements. Nevertheless, there is an inevitable trade-offs between the length of the speech samples and duration of the experiment. Using longer speech samples increases not only the duration of the experiment, but also the demands on raters. Another limitation of longer speech samples is that it is difficult to obtain long spontaneous speech from language learners if their oral proficiency is low. Due to all these concerns, one-minute of speech for each speaker composed of four 15-second snippets at different times in a spontaneous speech style (mainly from the questions/answers in the *Variety Show*) was randomly selected from the corpus. In order to examine whether speech performance by learners was improved as they progressed through the semesters, speech samples from different blocks of semesters were chosen.

For native speakers, all 11 Chinese instructors (9 females and 2 males) who fully acquired their L1, Mandarin, served as the baseline for comparing the results with heritage and English learners of Chinese. Four 15-second snippets were randomly selected from the database. For language learners, each semester had 15 recordings of class sessions, which were divided into 3 blocks. Two snippets were chosen from the first block, the beginning of the semester (the first five weeks) and two snippets were chosen from the last block, the end of the semester (the last five weeks). Between the blocks at the beginning and at the end of the semester, there as a four or five week gap. If a learner attended classes for more than one semester, 1 minute was chosen from each semester. Speech samples of 17 heritage speakers (5 females and 12 males) and 20 English learners of Chinese (5 females and 15 males) were randomly chosen based on the block design.

Seventeen heritage learners and 4 of the 20 English learners participated in the picture telling experiment as well. Three speech files of each subject in the clock telling task were randomly selected. As for simple picture description and complex picture description, three pictures were randomly chosen and then the speech files of these three pictures were used for each subject. All together, 398 speech files were chosen for analysis.

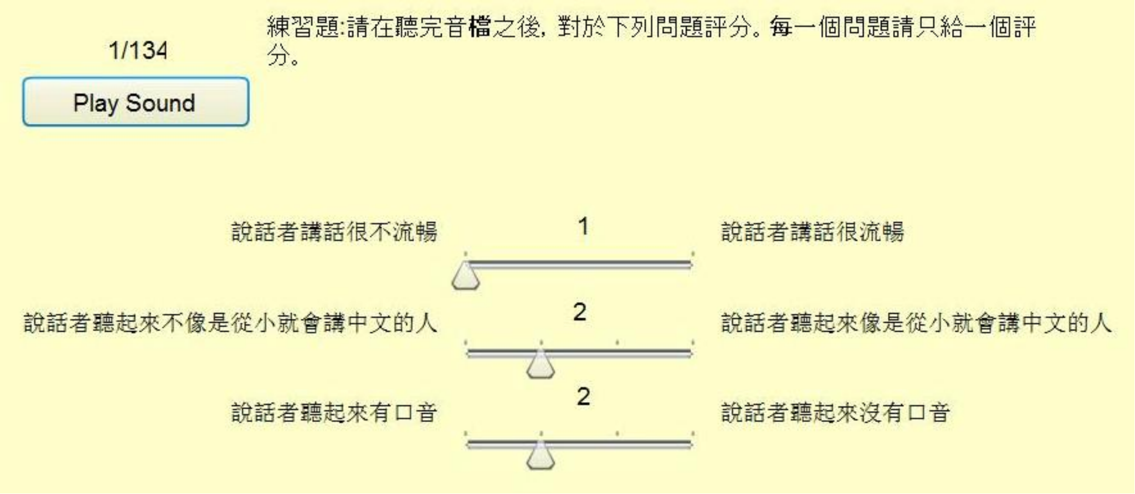
4.5.2 Perceptual Ratings

Forty-three native speakers of Mandarin in Taiwan rated all the 398 snippets. All the raters were untrained and linguistic naïve undergraduate students at National Changhua University of Education and National Chiao-Tung University. All raters reported normal hearing and were paid $20 (600 NTD) for participating.

The rating was conducted through a web interface built and maintained by ATLAS at UIUC. The whole experiment (398 snippets) were divided into 6 sessions due to the fact that the duration of evaluating 398 snippets at one time was too long. All the snippets were presented pseudo-randomly in each session. Eight questions were asked and they were presented in two pages where the audio files were played at least once for each page. At the beginning of each session, there was one training snippet. Sound files were auto-played when raters entered the question pages and it could be played as many times as needed. Once they submitted the answer and entered the next page, they were not allowed to go back to the previous page to change answers. The pictures of the simple picture description and complex picture description were shown with the corresponding speech files in a pop-up window.

The eight questions as shown in Figure 4.5 and Figure 4.6 represent speech performance about *fluency*, *nativeness*, *accentedness*, *disfluencies*, *pronunciation*, *grammar*, *vocabulary*, and *comprehensibility*. Each snippet was rated on a binary scale, 1 (not fluent) or 2 (very fluent), for the fluency rating and 4-point scale for the rest of the criteria. The design of binary fluency ratings is used for future development of automatic assessment system of fluency. The purpose is to set a threshold to classify speakers into two groups, 'fluent' or 'not fluent'. Nativeness based on speakers' identities should be a yes/no question. However, it is difficult to define whether heritage learners are native speakers or not. Moreover, it is interesting to see how listeners perceive a speaker as native or non-native speaker or somewhere between these two categories. Accentedness is a rating to measure the perceptual distance of speech between speakers and listeners, such as dialect accent and foreign accent. Dialect accent refers to differences of languages in different regions. For instance, people living in Hong Kong or Mainland China speak differently from people in Taiwan in terms of pronunciation, vocabulary, and grammar. When they speak a dialect in a particular region, their speech is perceived by speakers in other regions as having an accent. Likewise, when people learn a second or third language, they may speak differently from native speakers. This is called a foreign accent. Pronunciation, grammar and vocabulary are usually used as criteria in language testing or for language instructors to evaluate learner's performance in the classroom setting. Some questions, such as grammar and pronunciation, may lead to a more objective correct/incorrect judgment. Other criteria, such as accentedness, are subjective. Accentedness is an abstract concept and has no good or bad judgement. Grammar rating is used to measure the correctness of sentence structure. The vocabulary rating reflects the vocabulary size a speaker has and if he or she is able to select appropriate words. Disfluency is a quantifiers used widely to measure pauses, silence, self-corrections, repairs, repetition, etc in speech. This rating is used to gain a sense of disfluencies in perception. Comprehensibility evaluate whether a rater can comprehend the message delivered by speakers easily. Overall, the eight questions investigate how well the speech is perceived by native listeners.

A rater answered these questions for each speech file. The questions were listed on two pages as shown in Figure 4.5 and Figure 4.6. Positive scores are indicated by positive descriptions of the rating questions, presenting at the right side of each scale. Negative scores are indicated by negative descriptions of the rating questions, presenting at the left side of each scale.



練習題:請在聽完音檔之後，對於下列問題評分。每一個問題請只給一個評分。

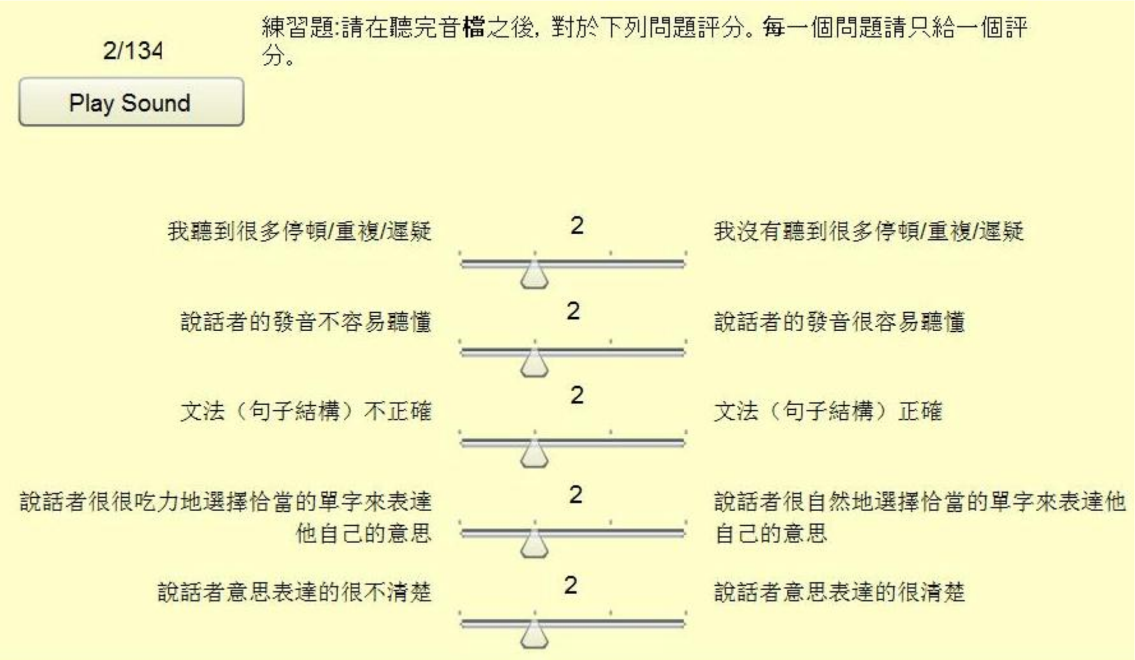
說話者講話很不流暢 1 說話者講話很流暢

說話者聽起來不像是從小就會講中文的人 2 說話者聽起來像是從小就會講中文的人

說話者聽起來有口音 2 說話者聽起來沒有口音

Figure 4.5: A sample of web page 1 with the first three rating questions.

- Page 1 with audio
  - Fluency: Is the speaker fluent or not? (1:not fluent; 2: fluent)
  - Nativeness: The speaker (doesn't) sound like a native Chinese speaker (1:not like a native speaker; 4: like a native speaker)
  - Accentedness: How accented is the speech? (1: accented; 4:no accent)



練習題:請在聽完音**檔**之後，對於下列問題評分。**每**一個問題請只給一個評分。

我聽到很多停頓/重複/遲疑	2	我沒有聽到很多停頓/重複/遲疑
說話者的發音不容易聽懂	2	說話者的發音很容易聽懂
文法（句子結構）不正確	2	文法（句子結構）正確
說話者很很吃力地選擇恰當的單字來表達他自己的意思	2	說話者很自然地選擇恰當的單字來表達他自己的意思
說話者意思表達的很不清楚	2	說話者意思表達的很清楚

Figure 4.6: A Sample of web page 2 with the rest five rating questions.

- Page 2 with audio
  - Disfluencies: I noticed a lot of pauses/repetitions/hesitations (1: a lot of disfluencies; 4: no disfluencies)
  - Pronunciation: The speaker's pronunciation was not easily understood (1:difficult to understand; 4: easy to understand)
  - Grammar: The speaker used sentence structure incorrectly (1: incorrect grammar; 4: correct grammar)
  - Vocabulary: The speaker had difficulty selecting appropriate vocabulary to express him/herself (1: difficulty selecting vocabulary; 4 no difficulty selecting vocabulary)
  - Comprehensibility: The speaker couldn't convey his/her intended message (1: unclear message; 4: clear message)

Figure 4.7 schematizes the dataset with information about the speakers, raters, task types and measurements. In sum, 398 snippets consisting of 236 snippets from the classroom setting and 162 snippets from the picture telling corpus were selected for perceptual rating. Only the 236 classroom snippets were submitted to acoustic analysis due to time constraints. In total, 48 speakers, including 11 native speakers, 17 heritage speakers and 20 English learners of Chinese from the database were chosen. Eight rating variables and 9 acoustic measures were examined.

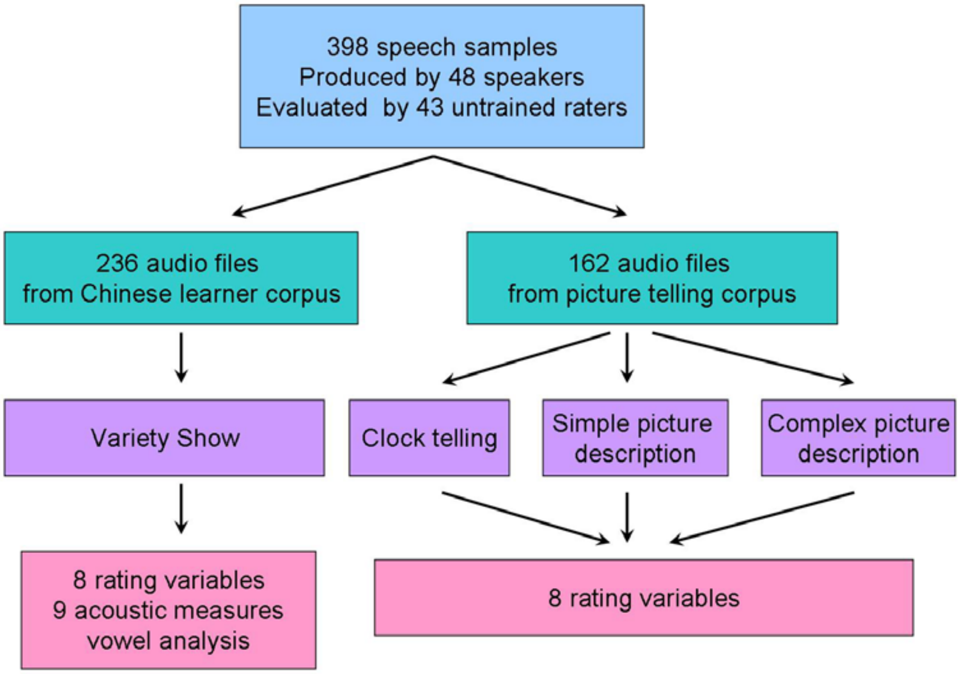


Figure 4.7: A schematic chart of the dataset for perceptual ratings and acoustic analysis.

4.6 Analyses

4.6.1 Rating Analysis

Learning Improvement

The mean rating scores for heritage and English learners at the beginning and the end of the semesters were presented in Table 4.1 and Table 4.2. A two-way repeated measures ANOVA was conducted to see whether there was any improvement of the rating scores between the beginning and the end of the semester with both Semester (2 levels) and 8 Rating variables (8 levels) as within-subjects factors. No significant difference was observed in the English learner group (F = 1.783, $p = .193$), while the heritage learner group showed slightly significant difference (F = 5.84, $p = 0.025 < 0.05$). Further analysis revealed that the effect size based on the mean rating scores of the heritage group is small (around 0.1 - 0.3). In addition, probably due to the fact that the time span between the beginning and the end of the semester was not long enough to observe the differences, the learning effect was not obvious. Thus, the data of different semester blocks were pulled together for further analysis.

Semester	Fluency	Native	Accent	Disfl.	Pron.	Grammar	Vocab.	Comp.
Begin	1.58 (0.22)	2.51 (0.57)	2.33 (0.45)	2.70 (0.52)	2.88 (0.52)	3.00 (0.49)	2.86 (0.58)	2.91 (0.59)
End	1.53 (0.25)	2.40 (0.59)	2.28 (0.44)	2.56 (0.47)	2.83 (0.54)	2.89 (0.44)	2.67 (0.55)	2.76 (0.57)
Effect size	-0.23	-0.19	-0.11	-0.14	-0.10	-0..22	-0.33	-0.25

Table 4.1: Mean rating scores for heritage learners at the beginning and at the end of the semester. Standard deviations are given in parentheses.

Semester	Fluency	Native	Accent	Disfl.	Pron.	Grammar	Vocab.	Comp.
Begin	1.31 (0.16)	1.66 (0.25)	1.62 (0.21)	2.15 (0.34)	2.08 (0.27)	2.44 (0.23)	2.17 (0.32)	2.17 (0.29)
End	1.26 (0.15)	1.62 (0.22)	1.62 (0.25)	2.04 (0.40)	1.99 (0.31)	2.33 (0.25)	2.02 (0.34)	2.04 (0.34)

Table 4.2: Mean rating scores for English learners at the beginning and the end of semesters. Standard deviations are given in parentheses.

Task Type Effect

The mean rating scores for each task type is shown in Table 4.3. Standard deviations are given in parentheses. A two-way repeated measures ANOVA was conducted to see whether there is a significant effect among task types on rating scores with Task Types (4 levels) and 8 Rating Variables (8 levels) as within-subjects factors. The result showed significant differences of task types (F = 13.438, $p < .001$). The interaction between task types and ratings is also significant (F = 13.997, $p < 0.001$). Pairwise comparisons revealed that the clock-telling task is significantly different from the other three task types ($p < 0.01$), whereas the other tasks (classroom data, simple picture description and complex picture description) have no significant differences with one another. Out of all of the rating scores in the tasks, clock telling held the highest. This is probably because the students were already in the third- and fourth- year of Chinese. A simple task is not able to differentiate the level of those variables when students reached a certain level of proficiency. This results in a ceiling effect. For instance, in the clock telling task, the response is very short (about 3 seconds) and the vocabulary is limited to numbers. The sentence structure of clock telling is fixed to certain formats. Thus, a simple task can be used to differentiate language learners at a lower proficiency level, but not intermediate or advanced learners. Figure 4.8

shows the rating variables in the horizontal axis and the scores in the vertical axis. The task types, classroom data, clock telling, simple picture description, and complex picture description are coded with the colors' navy, blue, yellow and red, respectively.

Tasks	Fluency	Native	Accent	Disfl.	Pron.	Grammar	Vocab.	Comp.
Clock	1.81 (0.14)	2.80 (0.43)	2.69 (0.37)	3.52 (0.33)	3.49 (0.26)	3.70 (0.16)	3.57 (0.22)	3.66 (0.22)
Simple	1.62 (0.23)	2.50 (0.60)	2.45 (0.51)	2.96 (0.45)	3.11 (0.43)	3.15 (0.32)	3.00 (0.44)	3.18 (0.42)
Complex	1.58 (0.23)	2.48 (0.64)	2.44 (0.52)	2.85 (0.53)	3.07 (0.49)	3.00 (0.38)	2.90 (0.51)	3.08 (0.47)
classroom	1.60 (0.24)	2.67 (0.66)	2.47 (0.46)	2.74 (0.52)	2.99 (0.55)	3.06 (0.48)	2.89 (0.60)	2.93 (0.59)

Table 4.3: Mean rating scores for task types. Standard deviations are given in parentheses.

One of the important observations here is that the ratings of the simple picture description and the complex picture description appear to pattern closely with the ratings of the classroom data. The duration of these two tasks is about 7 to 10 seconds and each snippet of the classroom data is 15 seconds. To closely examine whether the scores of the same learners are consistent across all task types, Figure 4.9 and Figure 4.10 show the plot of fluency as a function of accentedness and the plot of nativeness as a function of accentedness. These figures show that the construct of fluency is related to the construct of accentedness, such that learners who are more fluent tend to have less accent, and the constructs of nativeness and accentedness are in an almost perfect linear relationship. On the other hand, in Figure 4.9, the fluency scores of the same speakers are higher in the clock telling task than other tasks for students who were ranked low (e.g. speakers, LJZ, WSS, BR, FL, ZAJ. WSS is the most prominent example). This suggests

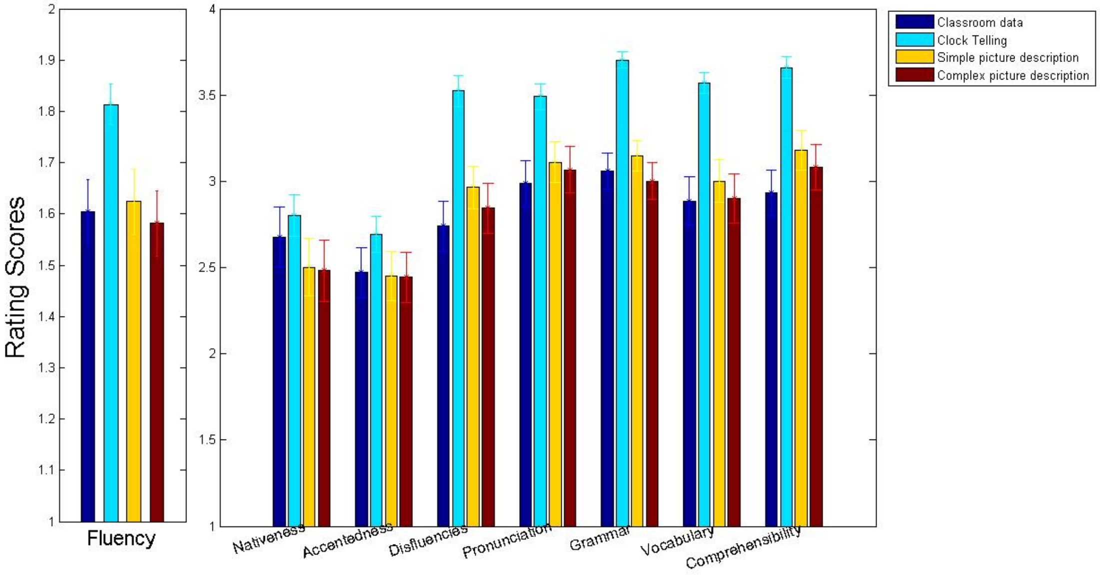


Figure 4.8: Bar plot of task types on rating scores.

that the measure of fluency varies, depending on task types. In Figure 4.10, the scores of nativeness and accentedness for the same speakers are consistent across all task types. This demonstrates that native listeners can get a global impression of language learners' nativeness and accentedness fast and consistently, which has been shown in the thin slice study of students' teaching evaluation (Ambady & Rosenthal, 1993).

Speaker Group Effect

Table 4.4 shows the average scores of each speaker group. The standard deviations are given in parentheses. A one-way repeated measures ANOVA was conducted on the classroom data to examine the differences of Speaker Groups (3 levels) as a between-subjects factor with 8 Rating Variables (8 levels) as a within-subjects factor. The results showed significant effects of speaker groups (F

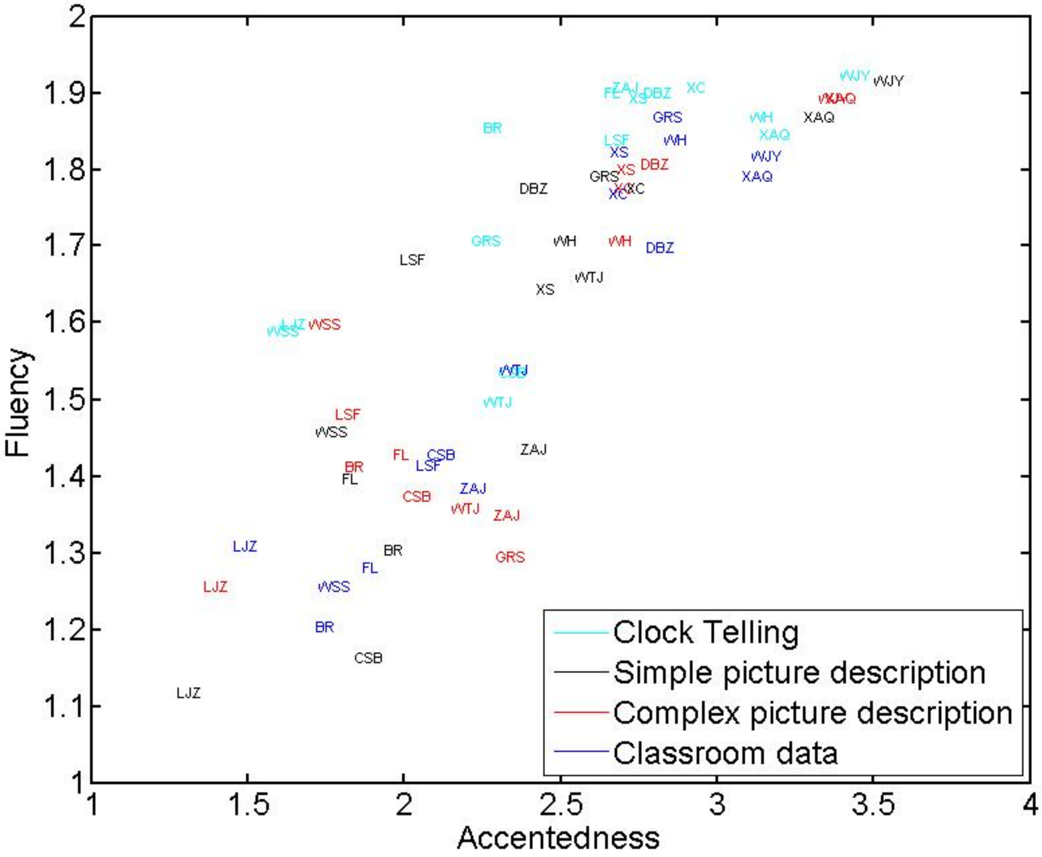


Figure 4.9: Correlation between fluency and accentedness in different task types

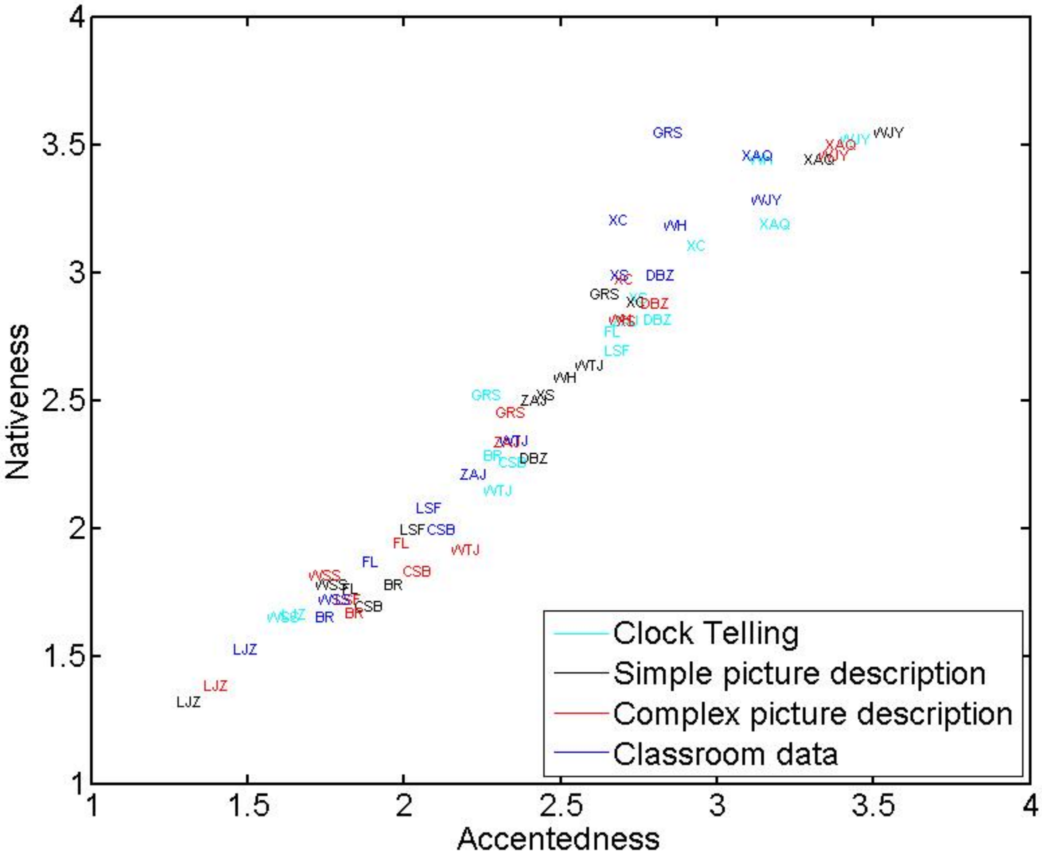


Figure 4.10: Correlation between nativeness and accentedness in different task types

Speakers	Fluency	Native	Accent	Disfl.	Pron.	Grammar	Vocab.	Comp.
Native	1.95 (0.04)	3.80 (0.10)	3.42 (0.33)	3.76 (0.12)	3.83 (0.12)	3.80 (0.09)	3.82 (0.09)	3.82 (0.13)
Heritage	1.60 (0.22)	2.59 (0.63)	2.38 (0.46)	2.72 (0.48)	2.93 (0.51)	3.02 (0.45)	2.86 (0.55)	2.90 (0.54)
English	1.31 (0.18)	1.69 (0.31)	1.66 (0.28)	2.16 (0.43)	2.11 (0.40)	2.43 (0.31)	2.17 (0.43)	2.18 (0.43)

Table 4.4: Mean rating scores for speaker groups. Standard deviations are given in parentheses.

= 4674.2, $p < 0.001$) and the interaction between speaker groups and ratings (F = 389.63, $p < 0.001$). *Post-hoc* tests using the Tukey HSD procedure revealed that the speaker groups were significantly different from one another. Figure 4.11 shows that native speakers have the highest scores, followed by heritage learners and then English learners of Chinese. The gaps between Mandarin native speakers and Chinese heritage learners are greater than those between heritage learners and English learners of Chinese. Figure 4.12 shows boxplots of 8 rating variables by different speaker groups. M represents Mandarin native speakers; H represents heritage speakers; and E represents English learners of Chinese. The distribution of the rating scores shows that the native speakers reach the ceiling of the scores, except accentedness ratings. Heritage learners and English learners of Chinese have an overlapping area. The heritage learners are of varying proficiency levels, spread across the scale. Some of them reach native-like oral performance, whereas others behave more like English learners who do not have the advantage of a Mandarin background. This might be due to the onset age of acquiring Mandarin, the exposure to Mandarin, and the daily usage of Mandarin which might all affect and contribute to a speaker's oral proficiency to different degrees. Only a few English learners received high scores like native speakers.

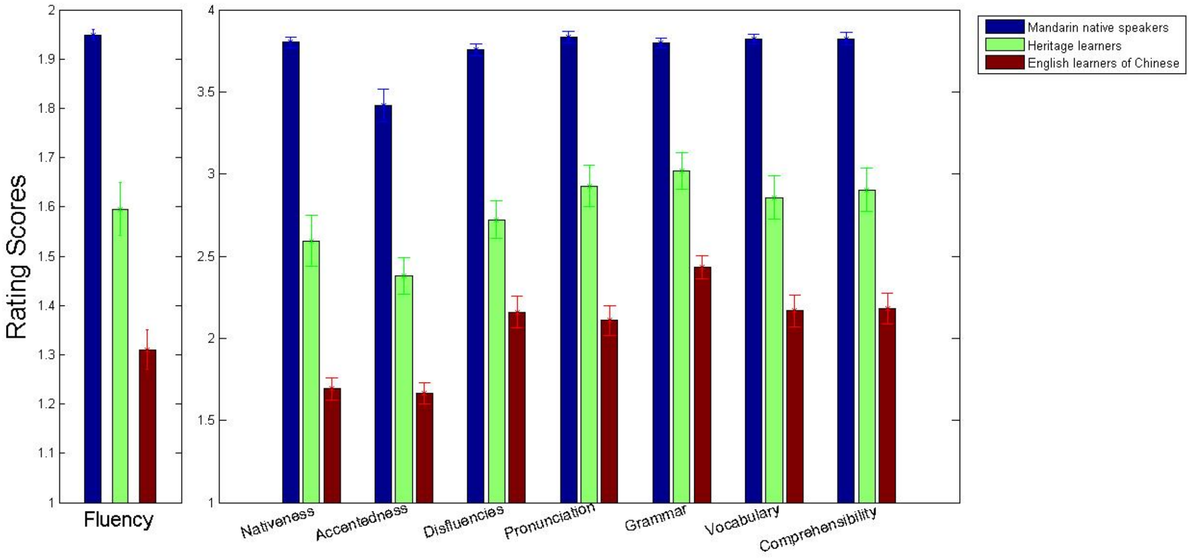


Figure 4.11: Bar plot of speakers groups on rating scores.

Table 4.5, Table 4.6, and Table 4.7 present the correlation matrix of rating variables in the classroom data by Mandarin native speakers, heritage learners, and English learners, respectively. Table 4.8 and Table 4.9 give the correlation matrix of the rating variables in the picture telling data by heritage learners and English learners, respectively. All correlations were significant at 0.001 level. These rating variables are highly correlated. This might be due to the possibility that all of these variables correlate with proficiency in Chinese. For example, both accentedness and grammar correlate with proficiency, and thus correlate with each other. In other words, not all of the variables are necessarily directly related, but because of their potential correlation with proficiency, they might all be good predictors of fluency and of foreign accent.

It is also observed that the strength of the relationships is the strongest in the heritage group than that of the native and the English learner groups in the classroom data. The magnitude of the correlations is similar between heritage and

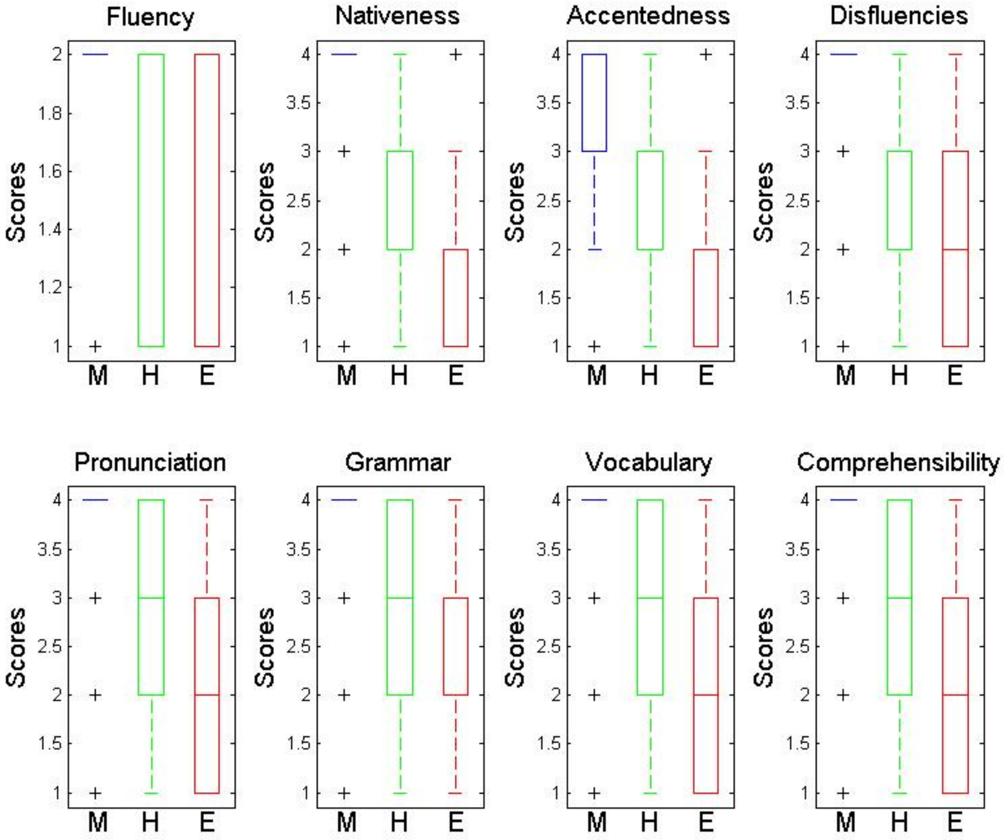


Figure 4.12: Boxplots of eight rating scores by speaker groups. M represents Mandarin native speakers; H represents heritage speakers; E represents English learners of Chinese.

English learner groups in the picture telling data. To determine whether the correlations are significantly different among speaker groups, a Fisher z' transformation of the correlation was performed and the difference was computed between different sized samples. The results reveal that the correlations of the heritage group do not significantly differ from the native speaker group but they do differ from the English learner group in the classroom data. The correlations of the English learner group are significantly different from that of the native group. As for the picture telling data, the correlations only show significant differences of ratings between nativeness and disfluency, vocabulary, and comprehensibility as well as ratings of accentedness and disfluency, pronunciation, vocabulary, and comprehensibility.

The findings of the correlation analysis are summarized as below.

- The ratings of fluency, disfluency and vocabulary highly correlate.
- The Nativeness rating correlates well with the accentedness rating in the heritage and English learner groups in both classroom and picture telling data. As for the native speakers, the nativeness rating correlates with the ratings of pronunciation, vocabulary, and comprehensibilty in the classroom data.
- The accentendess rating correlate with the pronunciation rating most.
- The ratings of pronunciation, grammar, vocabulary, and comprehensibilty highly correlate with one another.

corr (r)	Fluency	Native	Accent	Disfl.	Pron.	Grammar	Vocab.	Comp.
Fluency	1	0.86	0.54	0.72	0.84	0.79	0.87	0.84
Native		1	0.67	0.77	0.88	0.88	0.89	0.86
Accent			1	0.45	0.72	0.63	0.59	0.66
Disfl.				1	0.71	0.68	0.87	0.77
Pron.					1	0.92	0.88	0.94
Grammar						1	0.86	0.90
Vocab.							1	0.87
Comp.								1

Table 4.5: Correlation matrix of the rating variables of Mandarin native speakers in the classroom data. All correlations are significant at the 0.001 level

corr (r)	Fluency	Native	Accent	Disfl.	Pron.	Grammar	Vocab.	Comp.
Fluency	1	0.88	0.79	0.95	0.87	0.91	0.95	0.91
Native		1	0.94	0.85	0.91	0.87	0.88	0.86
Accent			1	0.79	0.91	0.83	0.83	0.83
Disfl.				1	0.86	0.91	0.96	0.92
Pron.					1	0.93	0.92	0.95
Grammar						1	0.96	0.97
Vocab.							1	0.97
Comp.								1

Table 4.6: Correlation matrix of the rating variables of heritage learners in the classroom data. All correlations are significant at the 0.001 level.

corr (r)	Fluency	Native	Accent	Disfl.	Pron.	Grammar	Vocab.	Comp.
Fluency	1	0.62	0.45	0.95	0.66	0.82	0.93	0.83
Nativ		1	0.93	0.56	0.80	0.69	0.68	0.68
Accent			1	0.37	0.79	0.58	0.51	0.58
Disfl.				1	0.59	0.82	0.96	0.84
Pron.					1	0.83	0.71	0.85
Grammar						1	0.90	0.95
Vocab.							1	0.92
Comp.								1

Table 4.7: Correlation matrix of the rating variables of English learners in the classroom data. All correlations are significant at the 0.001 level.

corr (r)	Fluency	Native	Accent	Disfl.	Pron.	Grammar	Vocab.	Comp.
Fluency	1	0.79	0.76	0.95	0.84	0.78	0.94	0.86
Native		1	0.96	0.75	0.89	0.68	0.80	0.78
Accent			1	0.69	0.89	0.64	0.75	0.75
Disfl.				1	0.77	0.81	0.95	0.83
Pron.					1	0.81	0.86	0.91
Grammar						1	0.91	0.94
Vocab.							1	0.94
Comp.								1

Table 4.8: Correlation matrix of the rating variables of heritage learners in the picture telling data. All correlations are significant at the 0.001 level.

corr (r)	Fluency	Native	Accent	Disfl.	Pron.	Grammar	Vocab.	Comp.
Fluency	1	0.80	0.75	0.95	0.88	0.88	0.96	0.88
Native		1	0.99	0.71	0.89	0.68	0.76	0.72
Accent			1	0.64	0.86	0.61	0.69	0.64
Disfl.				1	0.79	0.91	0.98	0.89
Pron.					1	0.81	0.86	089
Grammar						1	0.95	096
Vocab.							1	0.95
Comp.								1

Table 4.9: Correlation matrix of the rating variables of English learners in the picture telling data. All correlations are significant at the 0.001 level.

Among the rating variables, disfluency highly correlates to vocabulary. Vocabulary size is typically used to estimate lexical richness and diversity in corpus linguistics (Youmans, 1990) and language testing (Read, 2000). In addition, vocabulary size reflects a person's mental lexicon and productivity of language and speech. Bhat (Bhat, 2010) showed that the measure of lexical use (word types and word tokens) correlates well with fluency scores. Thus, word type and word count were measured in this study. Transcriptions of the classroom data were submitted to the Chinese word segmentation system developed by Chinese Knowledge and Information Processing (CKIP) at the Institute of Information Science, Academia

Sinica (Chen & Bai, 1998). The mean word type and word count are shown in Table 4.10. After word segmentation, the word type and word count of each snippet were submitted to a correlation analysis with disfluency scores for each speaker group, as given in Table 4.11. The word type represents how many distinct words were used in each 15-second snippet. The word count represents the total number of words used in each snippet. The results show that the correlation coefficients between disfluency scores and word types is 0.41, 0.72, 0.76 ($p < 0.01$) for Mandarin native speakers, heritage learners, and English learners, respectively. The correlation coefficiantes between disfluency rating and word counts is 0.19, 0.61, 0.73 for native speakers, heritage learners, and English learners, respectively. As expected, English learners had small word types (17.1) and word counts (25), followed by heritage learners (word types = 24.6; word counts = 34.6), and native speakers had the largest vocabulary size (word types = 33.3; word counts = 46.7). Since the native speakers are the instructors of Chinese language classes, their vocabulary size was expected to be sufficient for their oral performance. The difference in word counts could be that native speakers speak faster than learners, and thus their overall word counts should be higher. A Fisher z' transformation of the correlation was performed to compute the difference between two correlations (disfluency rating and word type, disfluency rating and word count). The results show that the correlations between native speakers and heritage/English learners was significant at the 0.05 level, while there was no significant correlation between heritage learners and English learners. This suggests that disfluencies may result from a lack of vocabulary in the L2 learners' speech planning.

Speaker Groups	Word type	Word Count
Native speakers	33.29	46.71
Heritage learners	24.55	34.61
English learners	17.12	24.94

Table 4.10: Mean word type and word count among speaker groups

Speaker Groups	Disflu/Word type (r)	Disfly/Word Count (r)
Native speakers	0.41**	0.19
Heritage learners	0.72**	0.64**
English learners	0.76**	0.73**

Table 4.11: Correlations between word type/word count and disfluency rating among speaker groups. $^{**}p < 0.01$

Correlation among Rating Variables

The correlation analysis revealed that all of the rating variables were highly correlated with one another, while the magnitude of correlations were different. Thus, the pattern of correlation analysis of the classroom data among speaker groups are examined in detail. Figure 4.13 depicts the correlation between nativeness and accentedness of the classroom data among speaker groups. The rating scores are the mean values of each snippet averaged from 43 raters. Note that accentedness correlates strongly with nativeness, especially in the learner group (native: r = 0.67; heritage: r = 0.94; English: r = 0.93, $p < 0.001$), implying that the higher the nativeness score, the less the accent is perceived. An interesting observation is that native speakers were perceived as native with scores ranging from 3.4 to 4, while their accent scores are between 3 to 4. To further examine the ratings of the native group, the mean scores of fluency, nativeness and accentedness were verified as shown in Table 4.12. The table shows that native speakers from Taiwan received higher accent scores (ranging from 3.4 to 3.88) than those from

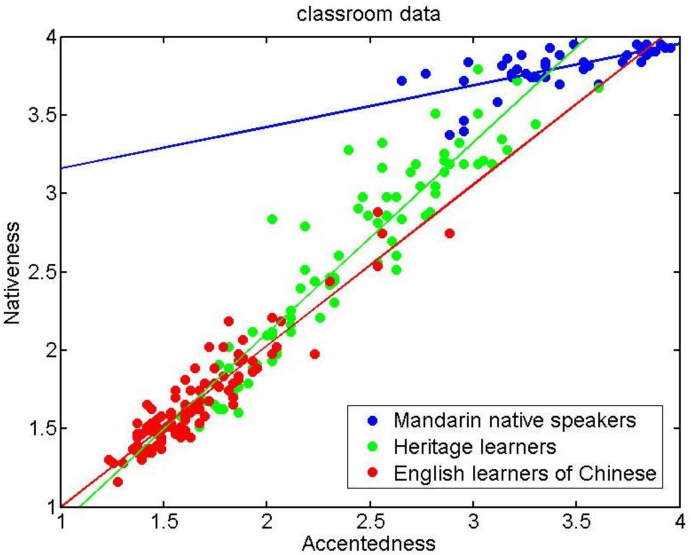


Figure 4.13: Correlation between Nativeness and Accentedness of the classroom data.

Mainland China (ranging from 2.9 to 3.4), while the fluency and nativeness ratings reach the ceiling in all native production. The reason is that the listeners are from Taiwan and they detected the different dialect accent of speech from Mainland China and ranked it lower, while they gave higher rankings to the speech that was closer to their own speech. The Taiwan speaker with the lowest ranking of accent has been in the U.S. for the longest time of all the native Chinese seaker, which may be a factor leading to differences in speech. This finding provides evidence for the concept of accent, which is defined as the perceptual distance of speech between listeners and speakers.

Figure 4.14 shows the correlation between fluency and accentedness of the classroom data among speaker groups (native: r = 0.54; heritage: r = 0.79; English: r = 0.45, $p < 0.001$). As we can see, the magnitude of the linear fit is not as tight as that between nativeness and accentedness. The heritage learner group shows the strongest relationship, while a still fairly large pool of learners received

Native Speakers	Gender	Originality	Fluency	Native	Accent
LHY	F	Taiwan	1.97	3.90	3.81
SJL	F	Taiwan	1.97	3.85	3.39
STJ	F	Taiwan	1.98	3.88	3.70
WZH	F	Taiwan	1.96	3.90	3.88
ZYX	M	Taiwan	1.98	3.86	3.80
GJ	F	Mainland China	1.89	3.66	3.10
GY	F	Mainland China	1.97	3.85	3.17
JY	F	Mainland China	1.95	3.77	3.09
WYJ	F	Mainland China	1.95	3.77	3.28
WS	F	Mainland China	1.95	3.77	3.44
LTL	M	Mainland China	1.87	3.59	2.94

Table 4.12: Mean rating scores of fluency, nativeness and accentedness in native group

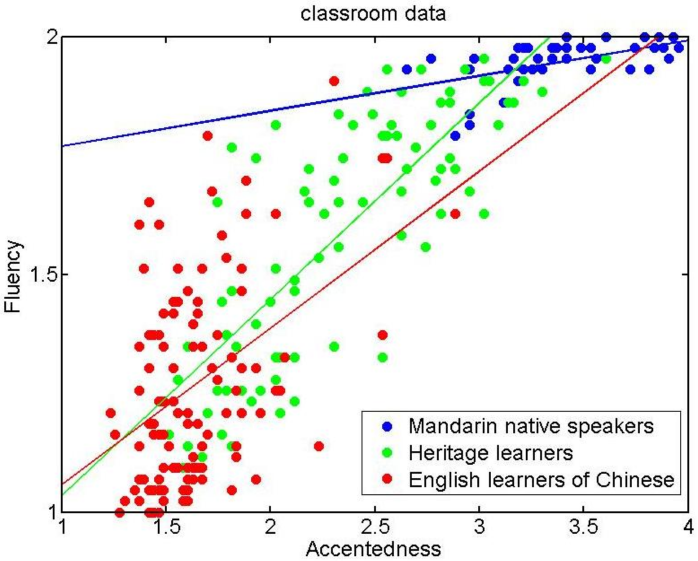


Figure 4.14: Correlation between fluency and accentedness of the classroom data.

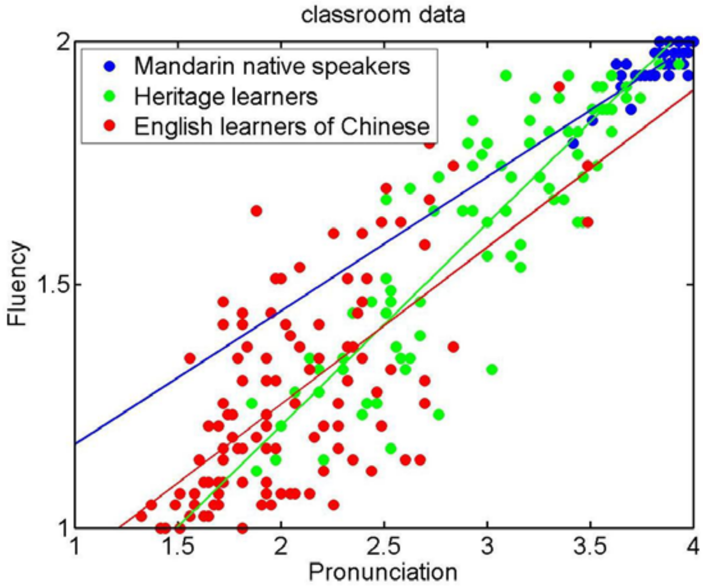


Figure 4.15: Correlation between fluency and pronunciation of the classroom data.

high fluency ratings (e.g. 1.65) with low accent ratings (lower than 2). Other learners received mild accent ratings (around 2.5) with relatively low fluency ratings (e.g. 1.3). The trend of the data distribution is not linear, indicating that a language learner can have a strong accent and still be judged as being fluent while some have mild accent and yet not be perceived as being fluent.

Figure 4.15 shows the correlation between fluency and pronunciation of the classroom data among speaker groups, in which the magnitude of the linear fit is not tight (native: r = 0.84; heritage: 0.87; English: 0.66, $p < 0.001$). In trend shows that the higher the fluency ratings, the higher the pronunciation scores. It is also observed that some English learners received high fluency ratings (above 1.5) with low pronunciation scores (around 2 points), while other English learners obtained low fluency ratings (near 1 point) with good pronunciation scores (between 2.5 to 3). There are some English learners who gained high fluency scores (near 2) and high pronunciation scores (between 3 to 3.5). This suggests that a

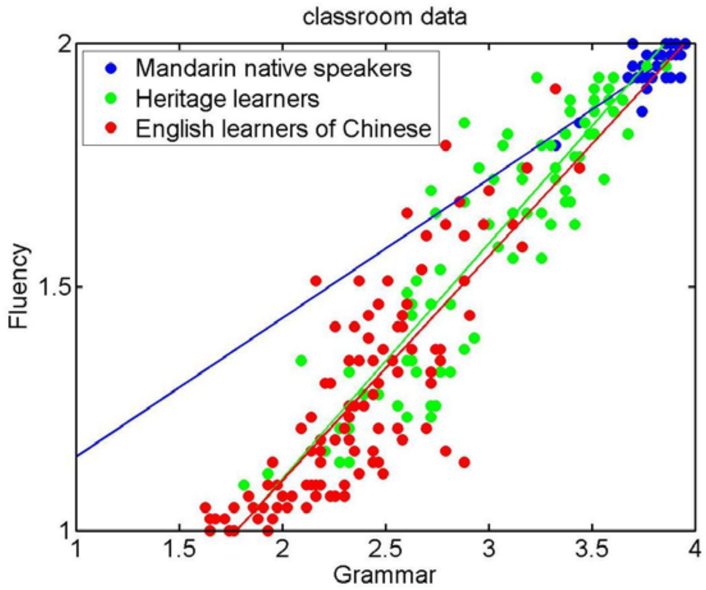


Figure 4.16: Correlation between fluency and grammar of the classroom data.

language learner can be fluent with poor pronunciation, and that one can have good pronunciation, but not be perceived as being fluent. On the bright side, it is possible for non-native speakers to have good pronunciation and speak fluently.

Figure 4.16 depicts the correlation between fluency and grammar of the classroom data among speaker groups (native: $r = 0.79$; heritage: $r= 0.91$; English: 0.93, $p < 0.001$). Most of the learner productions show higher grammar scores with high fluency ratings, while there are some learners who received high grammar scores (about 3), but low fluency ratings (near 1.2). A few cases have high fluency ratings (about 1.5), but low grammar scores (between 2 to 2.5). This demonstrates that some language learners can speak fluently with poor grammar or alternatively, speak disfluently with correct grammar.

Figure 4.17 presents the correlation between accentedness and pronunciation of the classroom data among speaker groups (native: $r = 0.88$; heritage: $r = 0.91$; English: $r = 0.88$, $p < 0.001$). The data distribution demonstrates that English

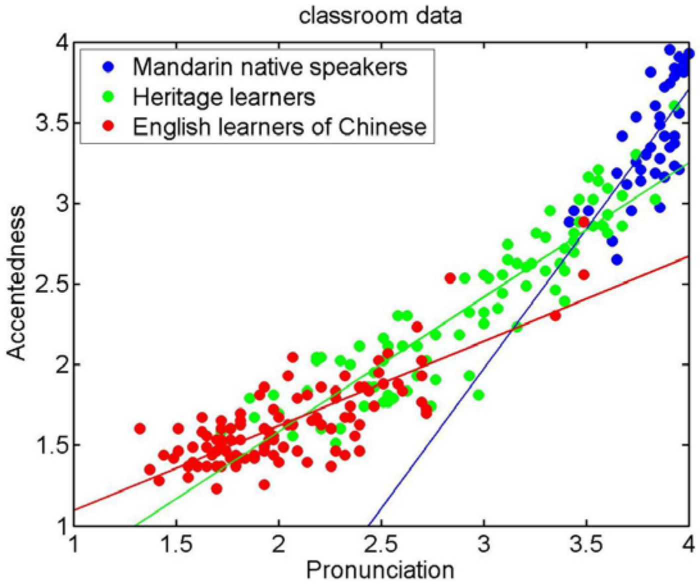


Figure 4.17: Correlation between Fluency and Accentedness of the classroom data.

learners are able to gain high pronunciation ratings (2.5 to 3), but the accentedness rating still remains low (1.5 to 2). Likewise, heritage learners are able to receive pronunciation ratings between 3.5 and 4, but their accentedness scores is between 3 and 3.5. This suggests that it is easier to improve pronunciation, than the impression of accent.

4.6.2 Principal Component Analysis

Since the rating variables highly correlated with one another, they were probably not independent. Moreover, this suggests that more than one variable might be measuring the same behaviour of the system. Therefore, a Principal Component Analysis (PCA) is applied for reducing dimensionality and revealing the internal structure of the data. PCA is one kind of exploratory factor analysis (EFA), which is a statistical model used to explore a reduced number of unobserved variables when there is no assumption or hypotheses about the construct of the measures. EFA decides the number of factors and chooses the extraction and rotation methods. PCA is a way which maximizes the variance and take into account all variability in the data. If we imagine that all of the rating scores were plotted together in a multi-dimensional space, the function of PCA is to rotate the whole dataset and find the best viewing angle to visualize the data at. Instead of using the original 8 rating variables to predict how well the speech is, PCA generates a new set of components based on the coefficients of the original variables. PCA creates the same number of principal components as the original variables (e.g. PCA creates 8 principal components based on the 8 rating variables), but usually the first two components are sufficient to explain the variance in the data. Each component is a linear combination of all the original variables with different weights, in which the component was formulated independently by the analysis.

In Figure 4.18, the first principal component is represented by the horizontal axis and the second principal component is represented by the vertical axis. The rating experiment yields 81,184 rating scores (8 questions x 43 raters x 236 snippets) of the classroom data and 55,728 rating scores (8 questions x 43 raters x 162 snippets) of the picture telling data. Each of the 8 variables is represented by a vector (the blue lines on the biplots). The length and direction of each vector indicate how each variables contributes to the two principal components, i.e., the projection of the vectors on the axes is the weight of original variables in the linear combination of the principal components. Vectors pointing in the same direction indicate the variables which are correlated. If the lines of the variables are long, it indicates a strong correlation between an axis and the corresponding variable. The direction of the vectors of the 8 rating variables are similar between the classroom and the picture telling data, suggesting that the variations are similar in

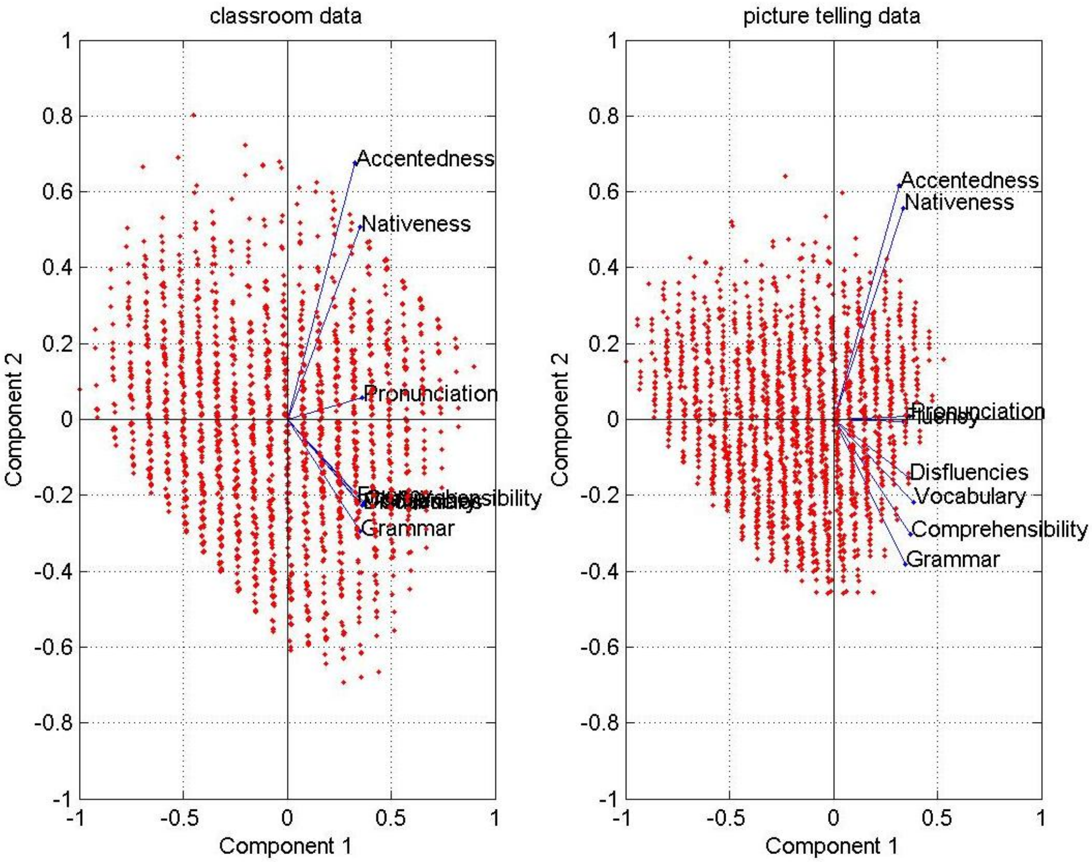


Figure 4.18: Principal component analysis of rating scores on the classroom and picture telling data.

these two datasets and that the results were not obtained by chance.

All of the rating variables contribute to the first principal component with similar weights, confirming what we have learned from the correlation matrix. Below are the equations of the first principal component (PC1) in two datasets, which are the linear combination of the eight variables. Each of the variables contribute to the weight, (coefficients) ranging from 0.32 to 0.38, for evaluating the speech in both classroom and picture telling data. The weight is the projection of each variable vector on the x axis (PC1).

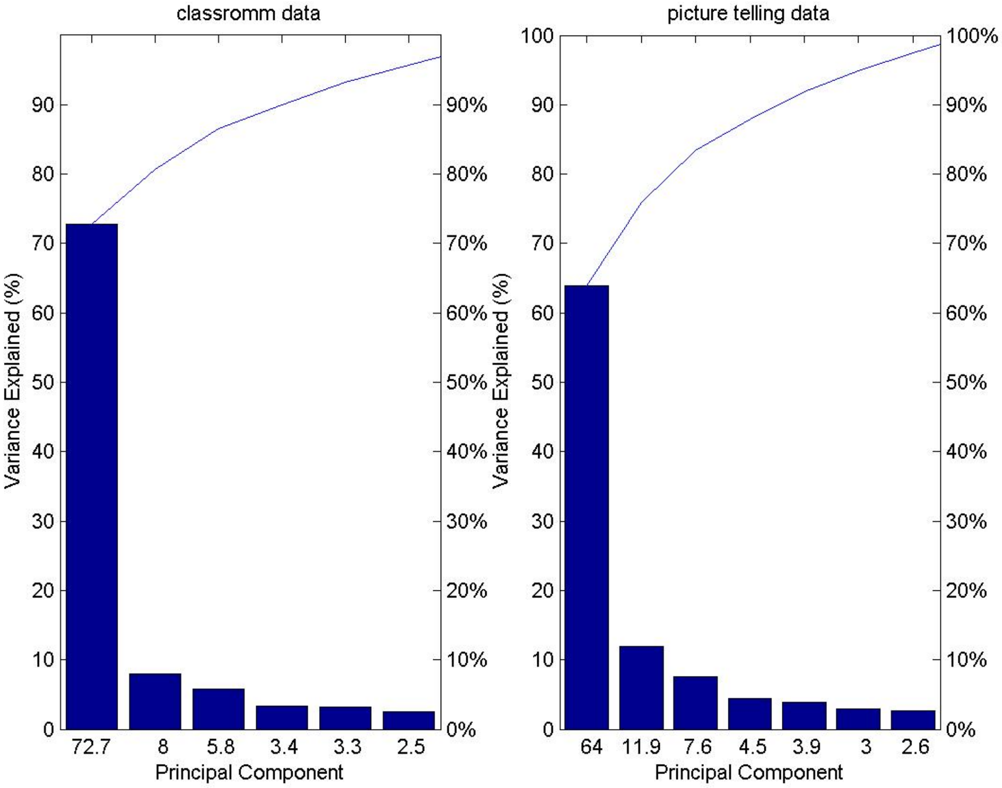


Figure 4.19: The percent of total variability explained by each component in the classroom and picture telling data.

1. classroom data: PC1 = 0.33*Fluency + 0.35*Native + 0.33*Accent + 0.36*Disflu + 0.36*Pron + 0.35*Grammar + 0.37*Vocab + 0.37*Comp
2. picture telling data: PC1 = 0.34*Fluency + 0.34*Native + 0.32*Accent + 0.36*Disflu + 0.37*Pron + 0.35*Grammar + 0.38*Vocab + 0.37*Comp

The second principal component is a dimension containing the sound-related factors of *Accentedness*, *Nativeness*, *Pronunciation* and the knowledge-related factors of *Fluency*, *Disfluency*, *Grammar*, *Vocabulary*, and *Comprehensibility*. Below are the equations of the second principal component (PC2) in the classroom and picture telling datasets. It appears that sound-related factors (positive coefficients/weight/projection on the y-axis) are a measure of how good the pronunciation is, while the knowledge factors (negative coefficients/weight/projection on the y-axis) is a measure of whether speakers can form speech and express ideas smoothly. Note that the lines for the sound-related factors and the knowledge-related factors are nearly perpendicular, indicating that they do not behave similarly. The weight contributing to PC2 is similar in both two datasets, where Accentedness has the heaviest weight > 0.6, followed by Nativeness > 0.5. Pronunciation has the least weight on PC2 (less than 0.1).

1. classroom data: PC2 = -0.21*Fluency + 0.5*Native + 0.68*Accent - 0.23*Disflu + 0.06*Pron - 0.3*Grammar - 0.23*Vocab - 0.22*Comp
2. picture telling data: PC2 = -0.006*Fluency + 0.56*Native + 0.62*Accent - 0.15*Disflu + 0.008*Pron - 0.38*Grammar - 0.21*Vocab - 0.3*Comp

Figure 4.19 shows that a clear break of the amount of the variance explained by each component is between the first and second components. The PC1 explains 72.7% and 64% of the variance in the classroom and picture telling data, respectively. The PC2 explains 8% of the variance in the classroom and 11.9% in picture telling data. Combining PC1 and PC2 can explain about 80% of the variance in the data. Even the third component is able to explain 5.8% and 7.6% of the variance, respectively, in the classroom and picture telling data, which yields significance at the 0.5 level.

For further understanding of the PCA results, the raw rating scores were imposed on the PCA dimensions, as shown from Figure 4.20 to Figure 4.27. The color bands in Figure 4.20 show the fluency scores of 1 in black and 2 in red. The color bands in the rest of the figures show the rating scores of 1, 2, 3, and 4 in black, red, green, and blue, respectively. The direction of the rating scores reinforces the relationship between the rating variables and the principal components, as presented in Figure 4.18. For instance, the direction of fluency ratings in the classroom data in Figure 4.20 reflects the direction of the vector ‘Fluency’ in Figure 4.18. The way to read the direction of the rating variables is to follow the score from low to high (1 to 4). Both nativeness in Figure 4.21 and accentedness in Figure 4.22 lie on the first quadrant (positive first and second principal components). The direction of fluency, disfluency, grammar, vocabulary, and comprehensibility all locate on the the fourth quadrant (positive first principal component and negative second principal component). Accentedness is perpendicular to comprehensibility, suggesting that they are fairly independent to each other.

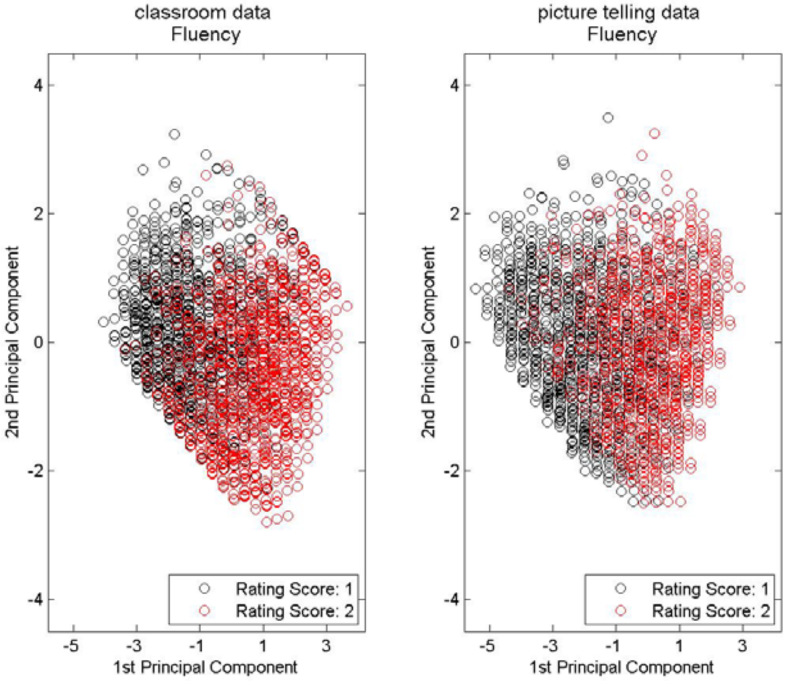


Figure 4.20: Principal component analysis of *Fluency* rating scores.

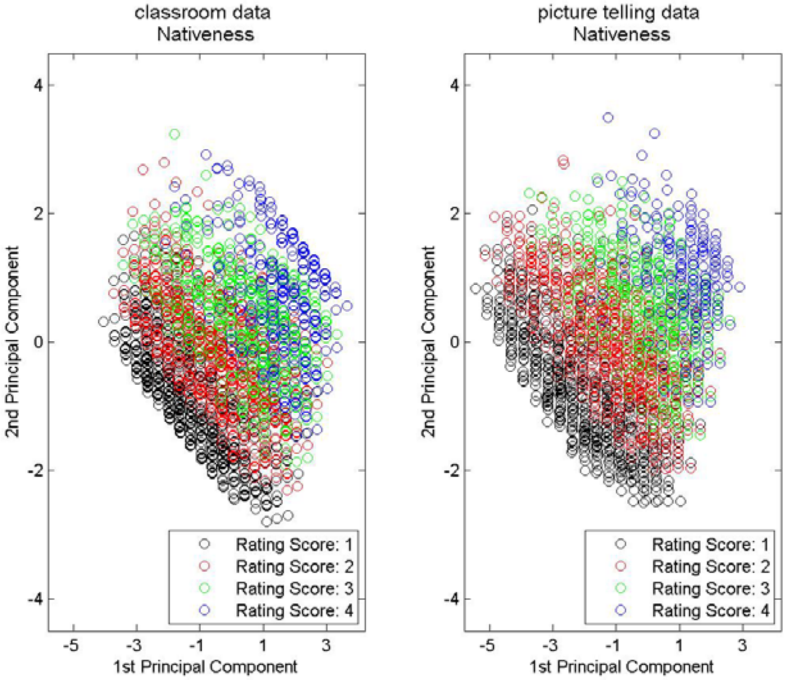


Figure 4.21: Principal component analysis of *Nativeness* rating scores.

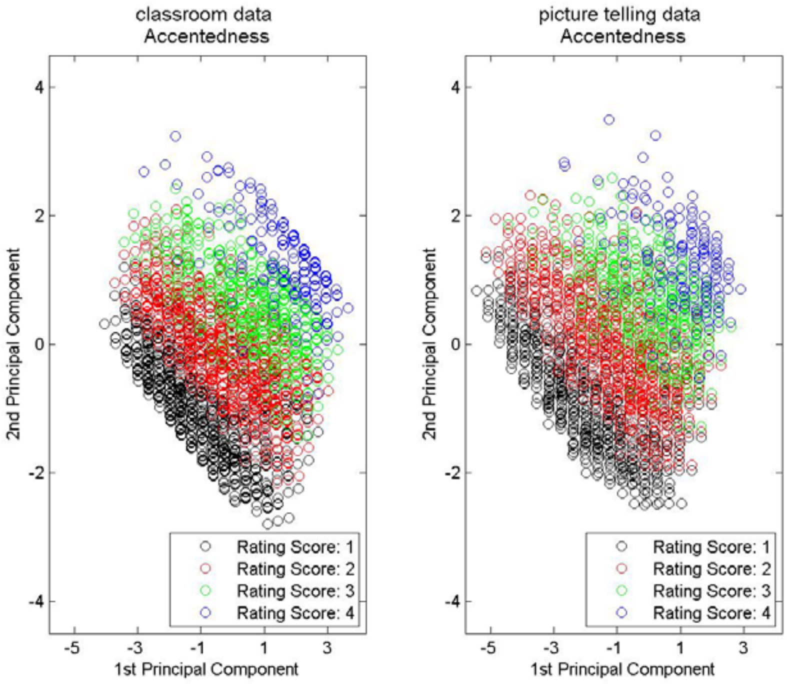


Figure 4.22: Principal component analysis of *Accentedness* rating scores.

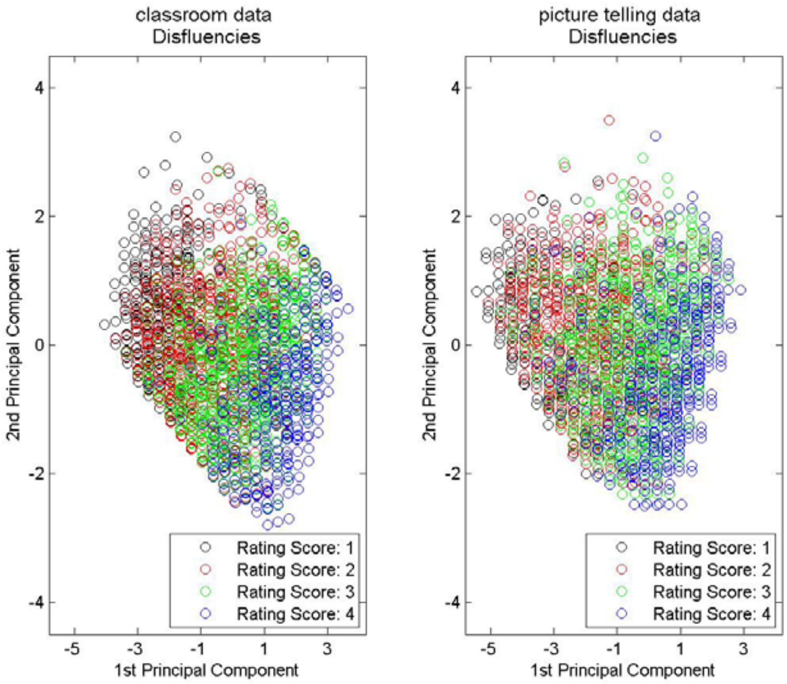


Figure 4.23: Principal component analysis of *Disfluencies* rating scores.

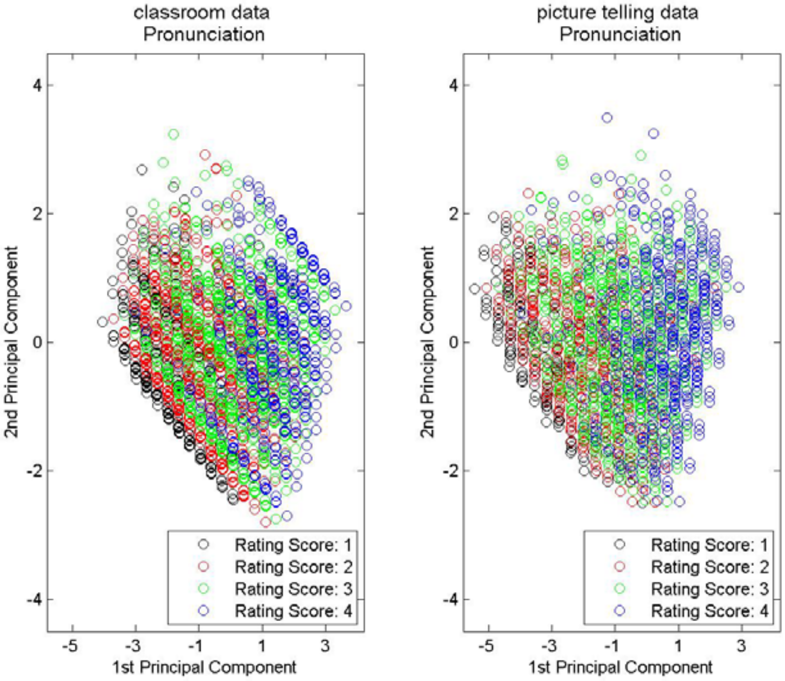


Figure 4.24: Principal component analysis of *Pronunciation* rating scores.

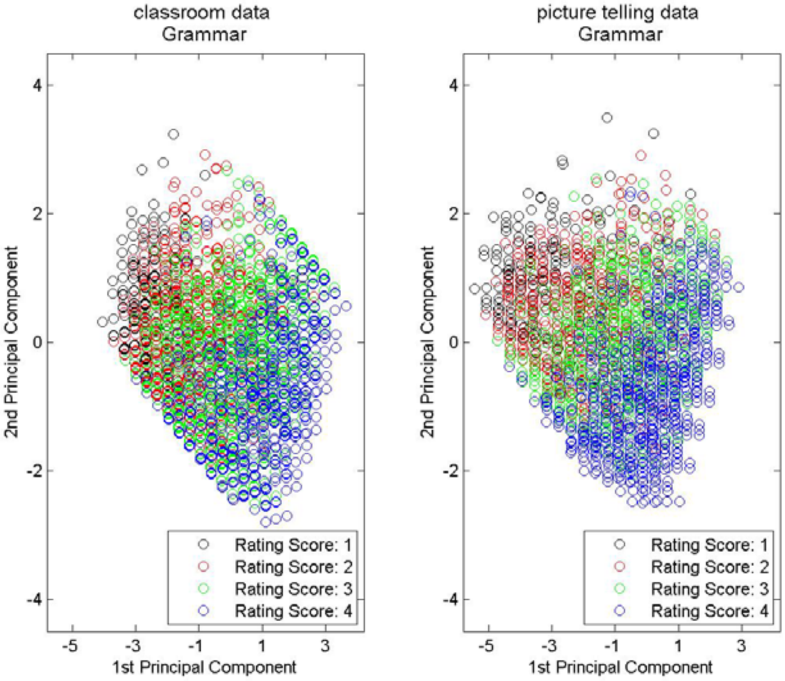


Figure 4.25: Principal component analysis of *Grammar* rating scores.

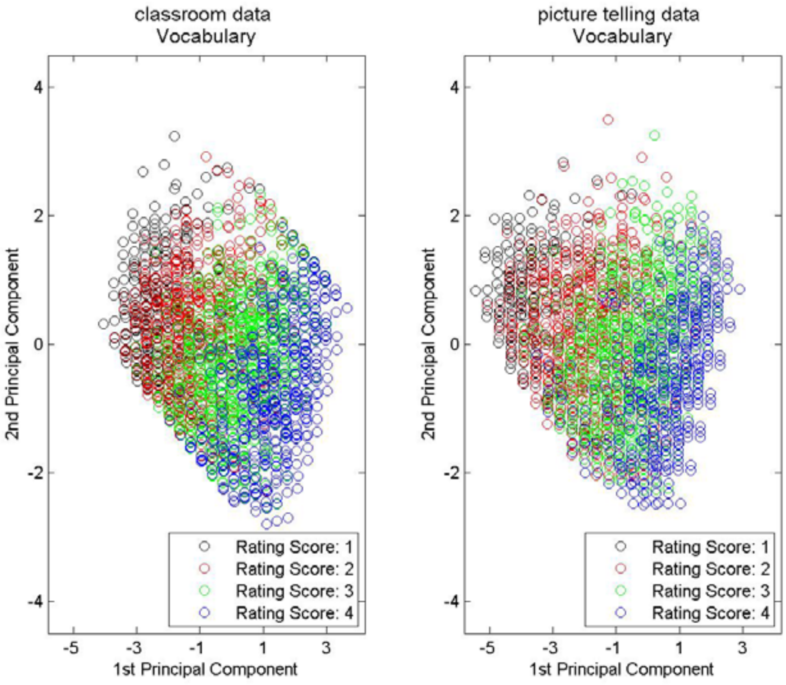


Figure 4.26: Principal component analysis of *Vocabulary* rating scores.

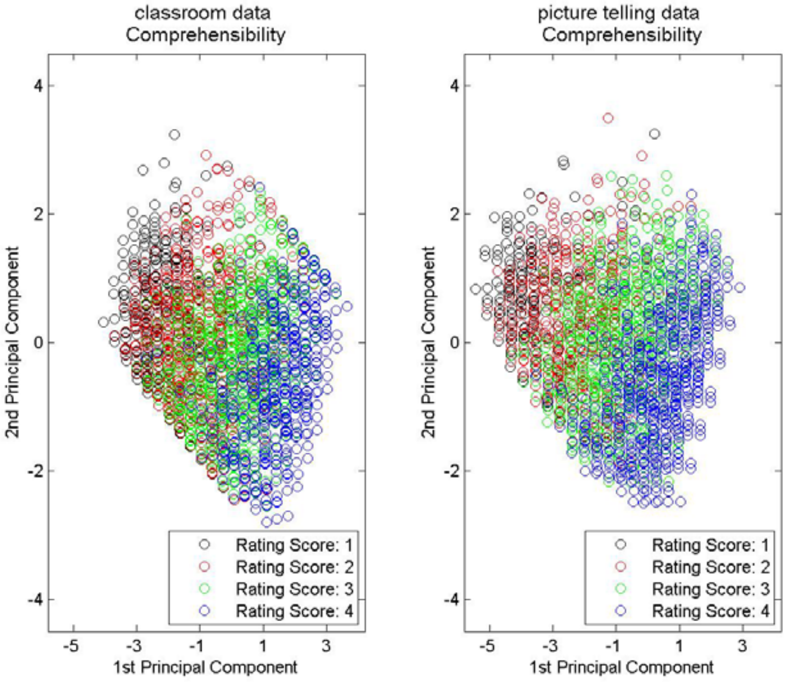


Figure 4.27: Principal component analysis of *Comprehensibility* rating scores.

4.6.3 Acoustic Analysis

Before running an acoustic analysis, phone labels were obtained for the 236 snippets in the classroom setting using the Penn Phonetics Lab Forced Aligner (P2FA) (Yuan & Liberman, 2008). In order to improve the alignment, I added a dictionary containing Zhuyin symbols, speakers codes corresponding to the name pronunciation used in speech, in addition to noise and disfluency transcriptions. The outcome of the automated phone segmentation was inspected and corrected manually by the author. With phone segmentation, vowel formants, duration values as well as following acoustic measures (Cucchiarini et al., 2000; Ramus et al., 1999) were automatically extracted for further analysis.

1. FPct: Number of filled pauses such as uh's and um's
2. FPdur: Duration of filled pauses
3. Articulation Rate (AR): Number of vowels (syllable nuclei) / utterance duration without silence
4. Rate of Speech (RS): Number of vowels (syllable nuclei) / utterance duration including silence
5. Phonation time ration (PTR): Utterance duration without silent pauses / utterance duration with silent pauses
6. Percentage Vowels (Pv): Duration of vowels (vocalic segments) / utterance duration without silence
7. Standard deviation of vowel duration (stdV)
8. F1: Average F1 of each snippet
9. F2: Average F2 of each snippet

The acoustic attributes, FPct, FPdur, AR, RS, PTR, Pv and stdV, measured the temporal properties of the speech produced by individual speakers in each snippet. Table 4.13 shows the mean values of the acoustic measures for speaker groups. The distribution of the acoustic attributes is presented in Figure 4.28, where English learners have the most FPs and longest duration of FPs, followed by heritage learners, and then native speakers. The AR, RS, and PTR all reveal that native speakers speak faster than heritage learners and English learners. Pv shows that native speakers have slightly longer vowel production than other speaker groups, and stdV shows more variation in vowel length.

Speaker groups	FP number	FPdur (seconds)	AR (syl/sec)	RS (syl/sec)	PTR (%)	Pv (%)	stdV
Native Mandarin	1.40	0.38	5.36	4.16	77	46	0.46
Heritage Learners	1.98	0.67	4.42	2.87	65	43	0.39
English Learners	3.70	1.48	3.28	1.86	57	42	0.36

Table 4.13: Mean values of acoustic measures for speaker groups.

Figure 4.29 shows a linear relationship between acoustic measures and fluency rating. Table 4.14 gives the R-squared values. The R-squared values indicate that RS, PTR and AR are able to explain, respectively, 68%, 51%, and 38% of the variance in the fluency rating. The frequency and duration of FPs have a negative relationship with fluency rating, meaning that the more FPs there are or the longer the FPs, the lower the fluency rating. Other acoustic measures related to speaking rates, such as AR, RS, PTR show a positive linear regression, indicating that the faster the speaking rate, the higher the fluency ratings. It is not surprising that native speakers have faster speaking rate than heritage and English learners and that native speakers received higher fluency scores, while the fluency rating of heritage and English learner speech varies greatly.

In Figure 4.30, the relationships between RS and ratings of fluency and accentedness within each speaker group were examined. The first row presents the

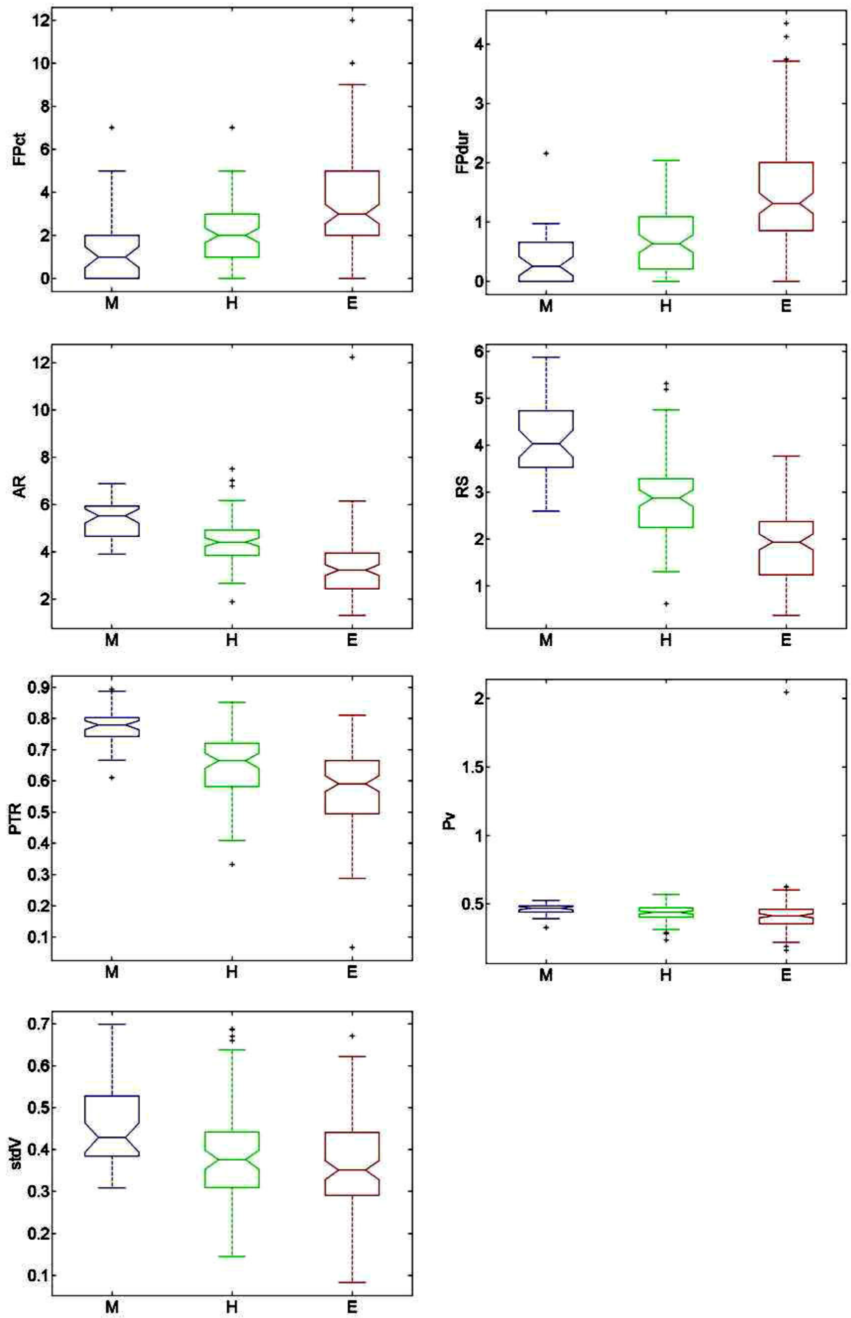


Figure 4.28: Boxplots of acoustic attributes.

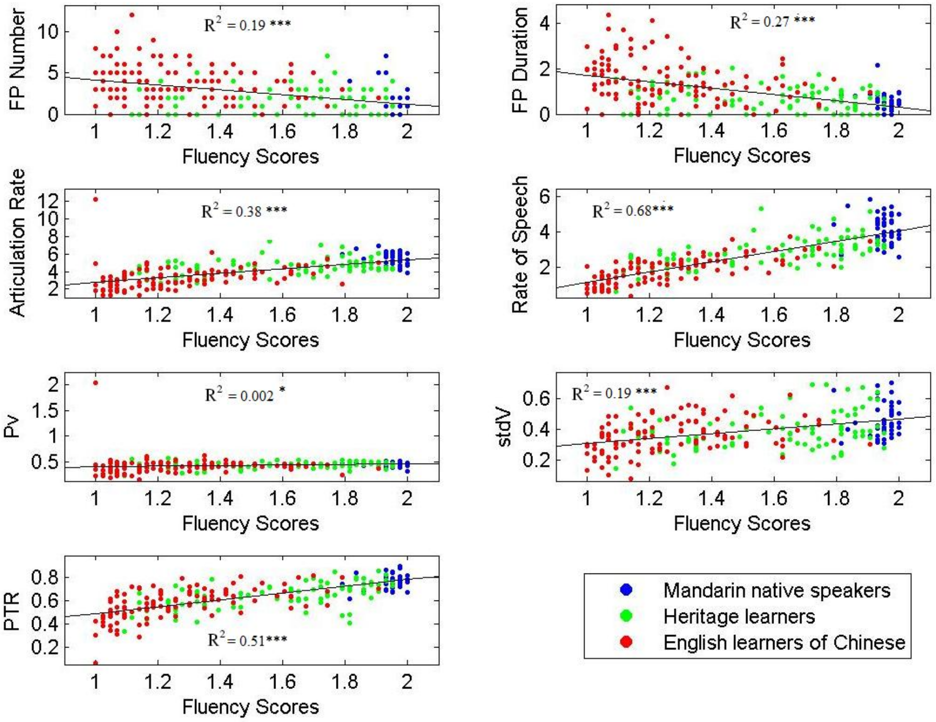


Figure 4.29: The relationship between acoustic measures and fluency rating (high scores indicates positive judgements). Dots in blue indicates native speakers; dots in green indicate heritage learners; and dots in red indicate English learners. $*p < 0.05$.

Acoustic measures	FPct	FPdur	AR	RS	Pv	stdV	PTR
R-squared	0.19*	0.27***	0.38***	0.68***	0.002*	0.19***	0.51***

Table 4.14: R-squared values of acoustic measures and rating scores. $***p < 0.001$, $*p < 0.05$

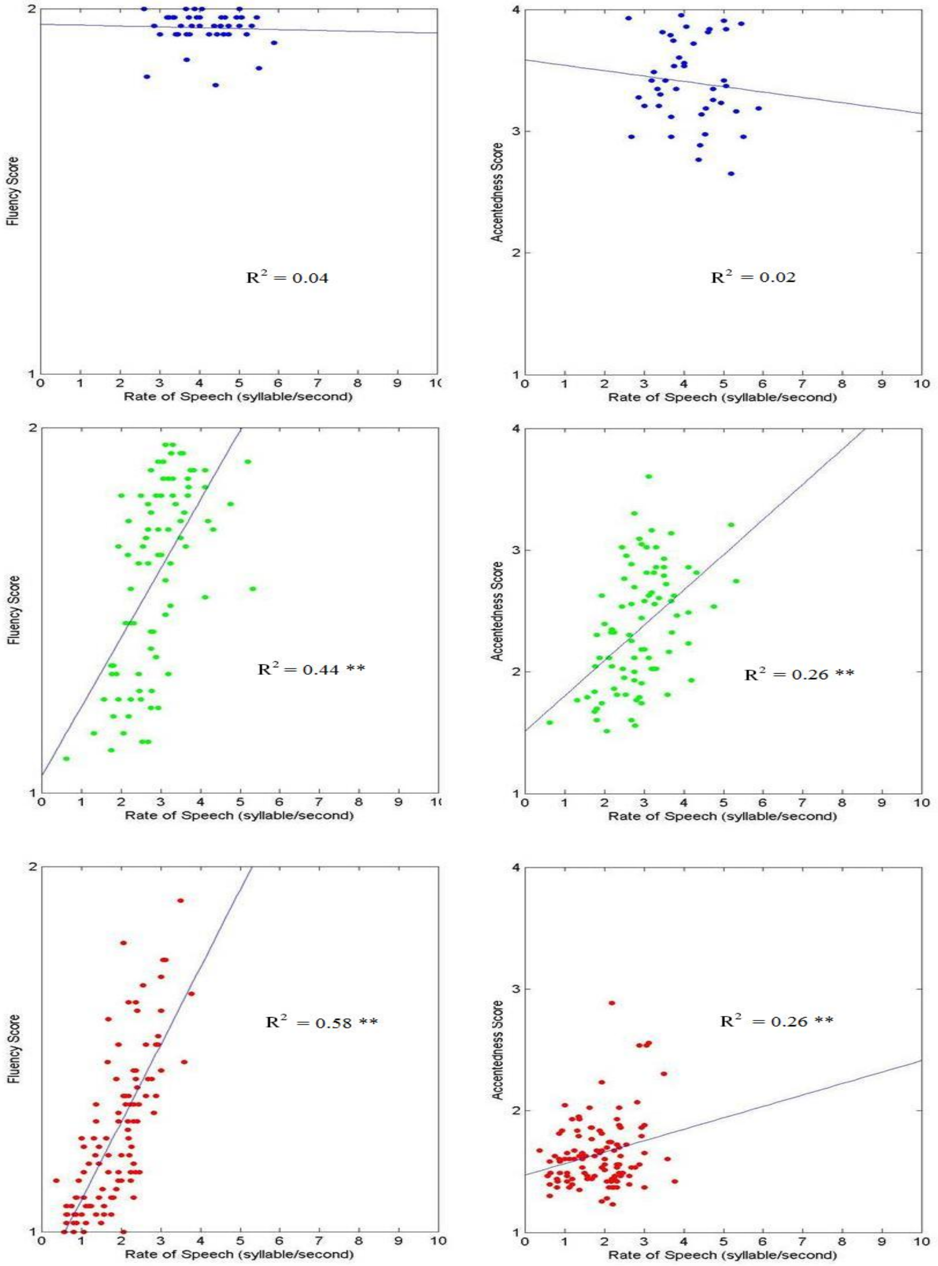


Figure 4.30: The relationship between rate of speech and the ratings of fluency and accentedness (high scores indicates positive judgements) within each speaker group. The first row in blue represents Mandarin native speakers; the second row in green represents heritage speakers; the last row in red represents English learners. $^{**}p < 0.01$

R-squared	RS/Fluency	RS/Accentedness
Native Mandarin	0.04	0.02
Heritage Learners	0.44**	0.26**
English Learners	0.58**	0.26**

Table 4.15: R-squared values of RS and rating scores; AR and rating scores. $^{**}p < .01$

plots of the native group; the second row presents the plots of the heritage group and the third row presents the plots of the English speaking learner group. R-squared values were given at Table 4.15. The r-squared values show that RS only accounts for less than 5% of the variance in fluency or accentedness ratings in the native group. Moreover, the linear fit of native groups is not significant, implying that RS does not affect the fluency rating for native speakers. The r-squared values revealed that RS accounted for 44% and 58% of the variance (p < 0.05) of perceived fluency in the heritage and English learner groups, respectively. The variance accounted for by RS in the fluency rating for the English learner group is more than that in the heritage learner group, suggesting that RS plays a more important role in English learners's fluency rating. Nevertheless, RS accounted for the same amount of variance (26%, $p < 0.05$) for the accentedness ratings in both heritage and English learner groups. Note that some learners have relatively high RS (e.g., above 4 syllables/per second), but their fluency or accentedness ratings are not higher than some speakers with slower RS. This needs further investigation to reveal other factors that influence the ratings.

In order to determine how acoustic measures load onto the PCA dimensions obtained from the rating scores, the first two principal components are expressed as linear functions of the acoustic measures. The PCA was conducted on the data. The mean rating score averaged 43 raters for each classroom snippet (236 snippets), which corresponded to the acoustic attributes extracted from each snippet. The projections of acoustic measures on the principal components are the coefficients between acoustics measures and rating components (x and y axes). As shown in Figure 4.31, RS(syllables/per second with silence in utterance), PTR (phonation time ration) or AR (syllable/ per second without silence in the utterance) is a powerful predictor of fluency. This result is along the lines of Cucchiarini et al. (2000, 2002). Lines pointing in the opposite direction indicate a negative correlation. Hence, the frequency and duration of FPs demonstrate a negative relationship with groups of temporal features and knowledge-related rating variables. With regard to detecting foreign accent, F1 and F2 are in the group of accentedness, nativeness, and pronunciation. The F2 of vowels is the most predictive factor among those acoustic attributes because it is close to the sound-related variables.

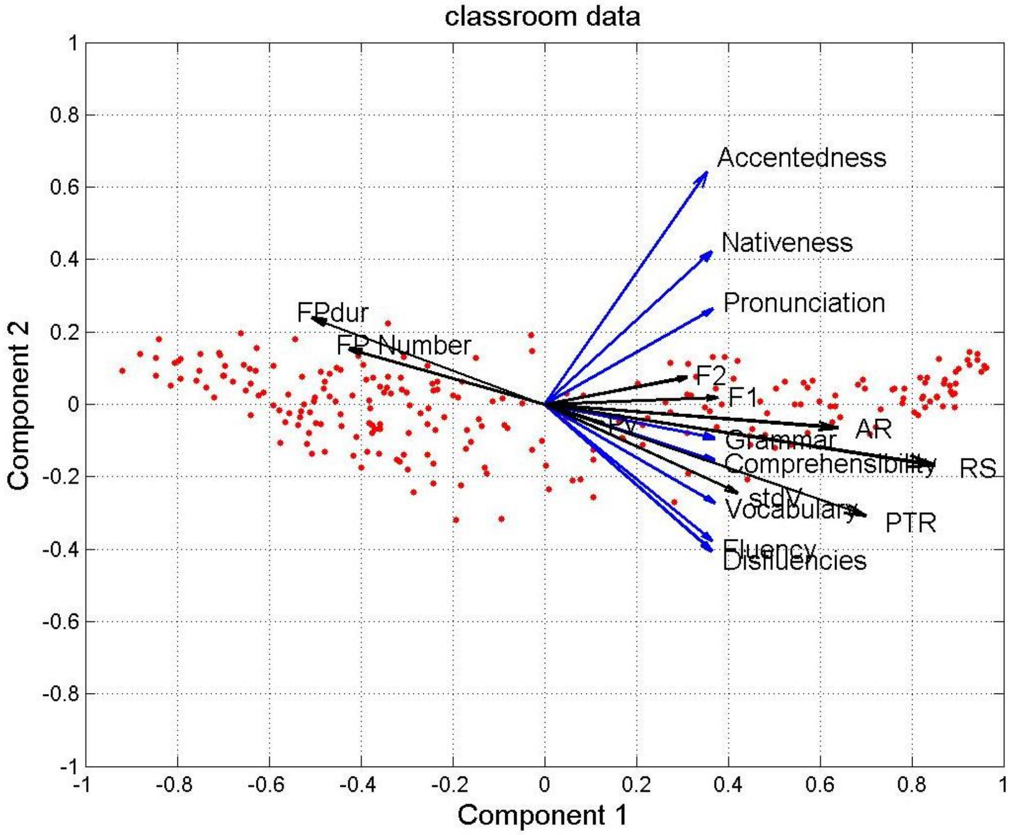


Figure 4.31: Principal component analysis of rating scores and acoustic attributes.

4.6.4 Vowel Analysis

Duration Data

The vowel duration data were submitted to a mixed design analysis of variance with Speaker Groups as a between-subjects factor and Vowel as a within-subjects factor. The analysis revealed significant main effects for Speaker Groups [F (1, 2) = 223.48, $p < 0.01$], and Vowel [F (1, 12) = 43.85, $p < 0.01$], as well as significant Speaker Groups X Vowel [F (15, 24) = 4.32, $p < 0.01$] interaction.

Vowel	Native Mandarin	Heritage learners	English learners
/i/	98.7 (61)	102.3 (65)	117.2 (81)
/ɨ/	94.4 (49)	102.2 (77)	123.0 (86)
/ɯ/	89.5 (58)	128.3 (94)	175.7 (145)
/u/	93.0 (59)	93.4 (61)	130.7 (88)
/y/	117.9 (42)	151.4 (96)	186.5 (138)
/e/	66.4 (34)	72.9 (40)	90.0 (53)
/ɛ/	79.0 (42)	90.4 (53)	109.6 (64)
/ə/	80.6 (61)	95.0 (74)	128.5 (106)
/o/	66.1 (35)	74.2 (45)	97.2 (53)
/ɔ/	90.8 (61)	104.5 (74)	145.6 (113)
/a/	92.3 (48)	100.3 (61)	116.8 (60)
/ɑ/	69.7 (31)	76.8 (35)	93.3 (43)

Table 4.16: Mean vowel duration (in msec) for speaker groups of Mandarin native speakers, heritage and English learners. Standard deviations are given in parentheses.

The mean durations of vowels (in msec) for the three groups are given in Table 4.16. English learners have the longest vowel duration, followed by heritage speakers. English learners might have the longest vowel duration due to their slower speaking rate among three speaker groups. *Post-hoc* tests using the Tukey HSD procedure revealed that vowel duration of [i, ɯ, u, ɛ, ə, o, ɔ, ɑ] by English learners was significantly longer than that of native speakers. English learners had significantly longer vowel duration for [u, ɔ, ɑ] than heritage learners did. Heritage learners produced [ɯ, ə] with significantly greater duration than the native speakers did, indicating that most of the vowel duration of heritage speakers was similar to that by native speakers. The comparison of vowel duration by speaker groups was shown in Figure 4.32.

Vowel duration and fluency both share temporal information in speech. Hence, vowel duration and fluency scores were submitted to a linear regression analysis. Figure 4.33 showed the relationship between vowel duration and fluency ratings with three speaker groups. Correlations are given in Table 4.17. It is observed that the longer the vowel duration, the lower the fluency rating. A question which arises here is whether the longer vowel durations might be due to a slower speaking rate among the English learners, while there is no such tendency in the native speaker group. To answer this question, the relationship between fluency ratings and vowel duration in each speaker group were submitted to a linear regression analysis. Figure 4.34, Figure 4.35 and Figure 4.36 present the correlation between fluency scores and vowel duration of the native Mandarin group, heritage learners, and English learners, respectively. The correlation coefficients for each group were given in Table 4.18. Indeed, the tendency observed here does not hold true for the native speaker group. Only the vowels [ɯ] and [o] showed a significant positive relationship between the rating scores and vowel duration ($p < .05$). In the heritage speaker group, vowels [u, e, o, ɔ] revealed a significant negative correlation between duration and rating scores. In the English speaking learner group, except for the vowels [ɨ, y, ɛ, a], there was a tendency that the longer the duration, the lower the scores. Therefore, the negative relationship between vowel duration and fluency ratings were contributed by L2 learners. Since the speaking

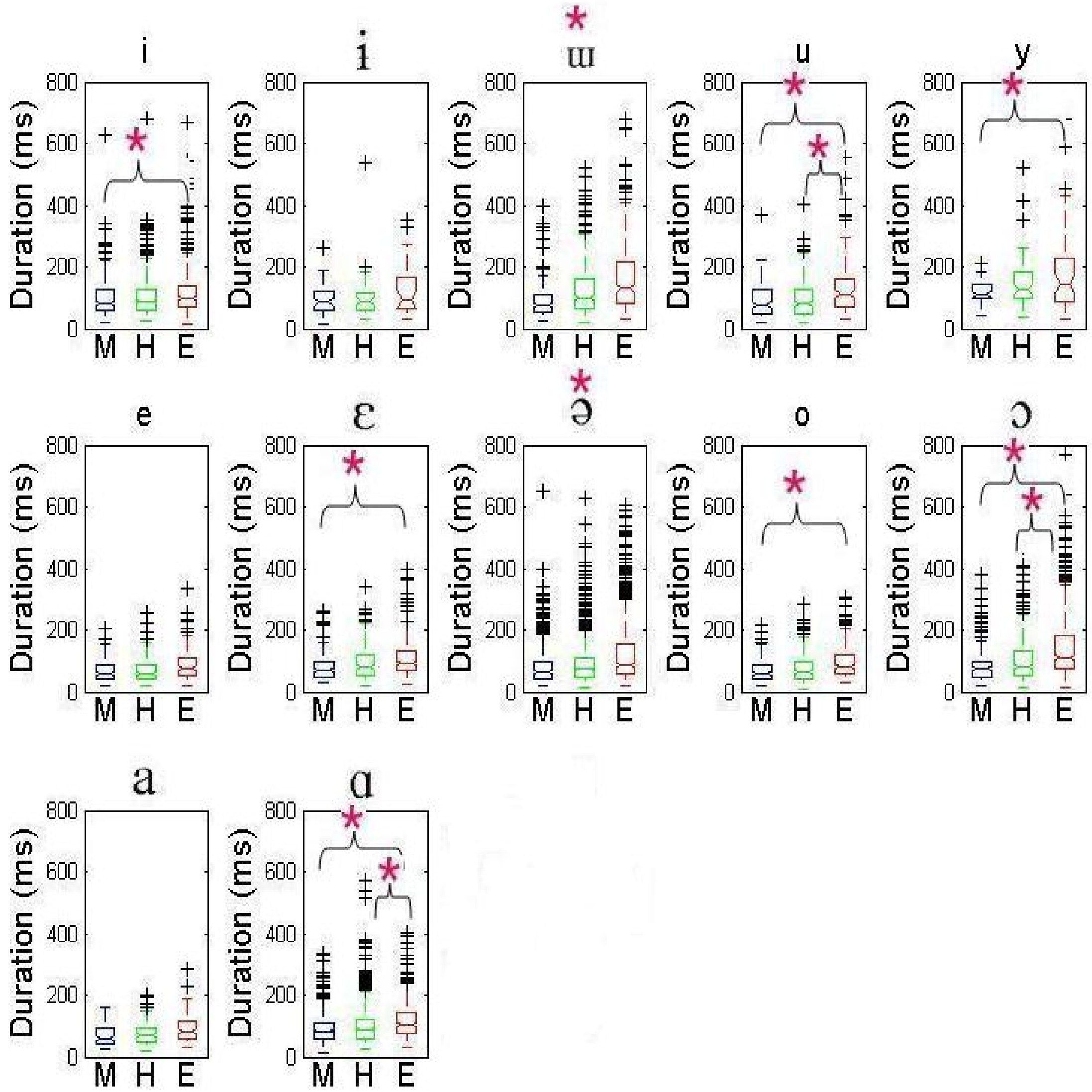


Figure 4.32: Boxplots of Vowel durations by speaker groups. M represents Mandarin native speakers; H represents heritage learners; E represents English learners of Chinese. The * on the vowel symbols indicates the difference of vowel duration among three groups is significant. $*p < 0.05$.

Vowel	Three groups
/i/	-0.13***
/ɨ/	-0.27**
/ɯ/	-0.29***
/u/	-0.27***
/y/	-0.22**
/e/	-0.26***
/ɛ/	-0.21***
/ə/	-0.24***
/o/	-0.26***
/ɔ/	-0.28***
/a/	-0.16***
/ɑ/	-0.24***

Table 4.17: Correlation coefficients (r) between vowel duration and fluency ratings of all speaker groups together as shown in Figure 4.33. $^{**}p < .01$, $^{***}p < .001$

rates of heritage and English learners were slower than native speakers, the longer vowel duration produced by L2 learners incorporates these properties of speaking rates. Another possible explanation is prolongation of speech sounds, which is a strategy to maintain speech flow (T.-L. Lee et al., n.d.). It is speculated that L2 learners might prolong words to maintain speech flow. Thus, the vowel duration produced by L2 learners was longer that that by native speakers.

Vowel	Native Mandarin	Heritage learners	English learners
/i/	0.02	-0.03	-0.15**
/ɨ/	0.10	-0.28	-0.28
/ɯ/	0.17*	-0.05	-0.25***
/u/	0.18	-0.19**	-0.22***
/y/	-0.07	0.04	-0.12
/e/	0.12	-0.16*	-0.19*
/ɛ/	-0.05	0.00	-0.10
/ə/	-0.02	-0.07	-0.18***
/o/	0.16*	-0.12*	-0.15*
/ɔ/	-0.12	-0.19**	-0.16***
/a/	0.06	-0.02	-0.16***
/ɑ/	0.19	-0.13	-0.05

Table 4.18: Correlation coefficients (r) between vowel duration and fluency ratings of each speaker group as shown in Figure 4.34, Figure 4.35 and Figure 4.36. $*p < .05$, $**p < .01$, $***p < .001$

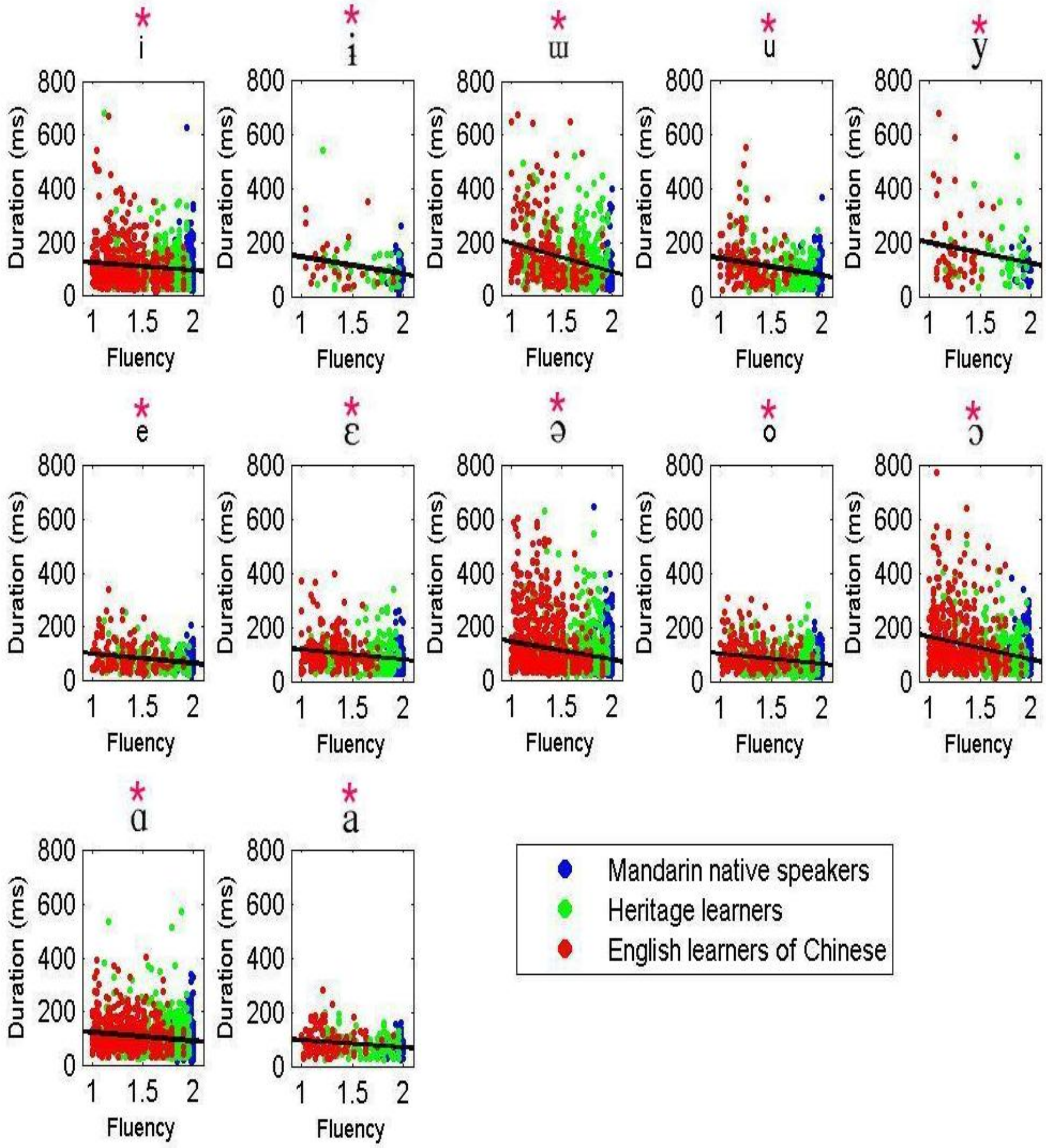


Figure 4.33: Correlation between vowel duration and fluency rating. Speaker groups are color-coded. Blue indicates Mandarin native speakers; green indicates heritage learners and red indicates English learners of Chinese. $*p < 0.05$.

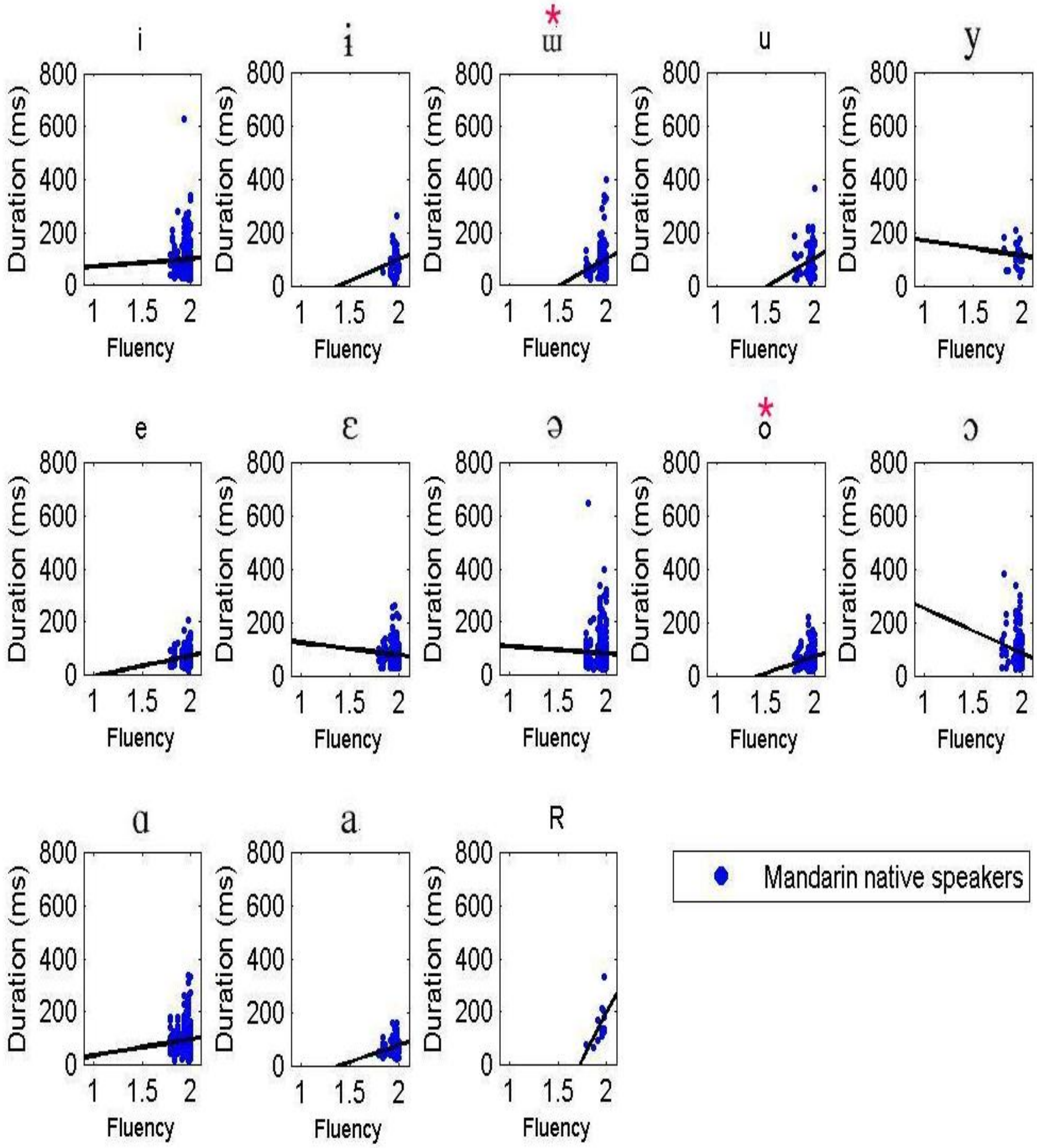


Figure 4.34: Correlation between vowel duration and fluency rating for Mandarin native speakers. $*p < 0.05$.

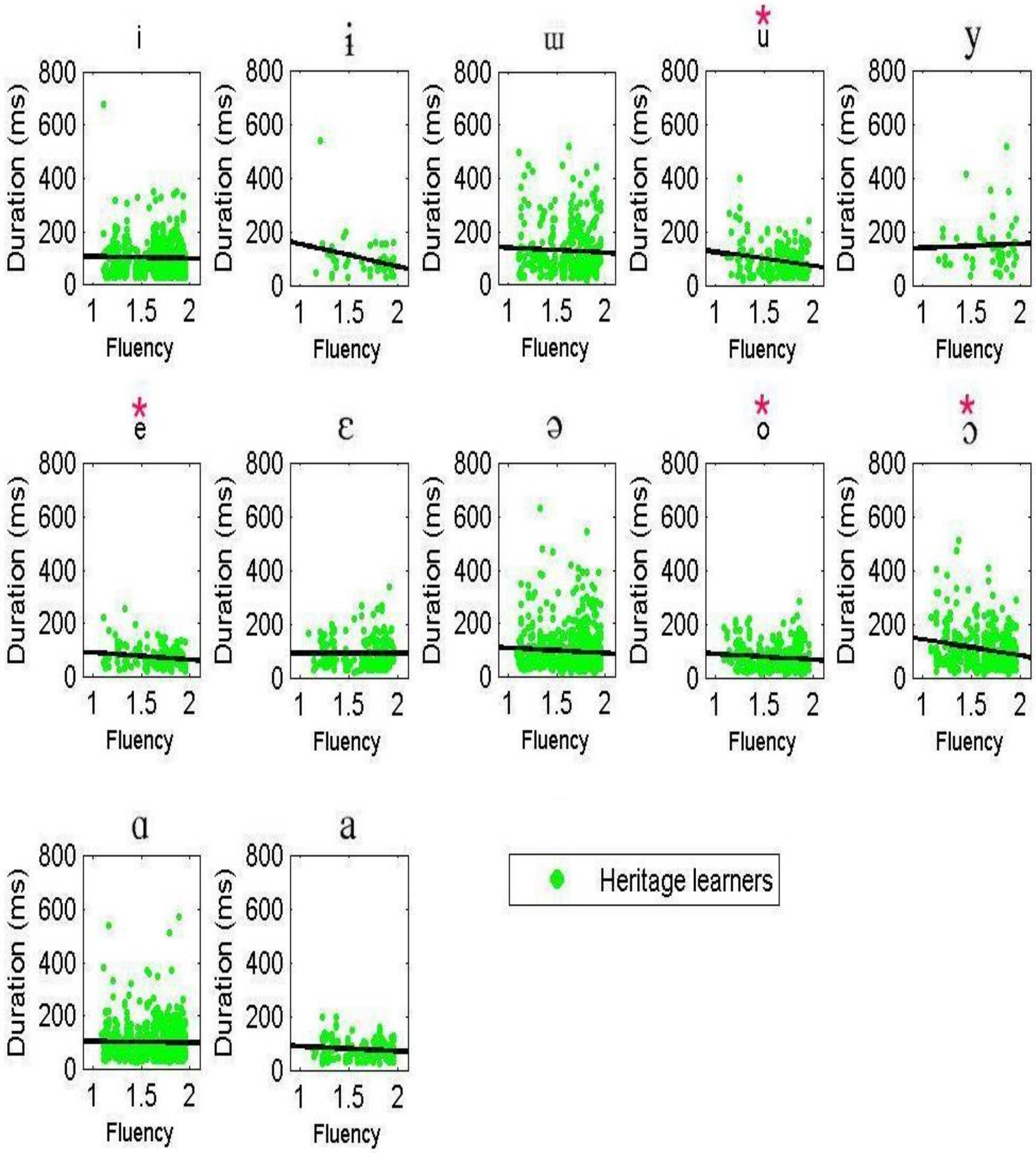


Figure 4.35: Correlation between vowel duration and fluency rating for heritage learners. $*p < 0.05$.

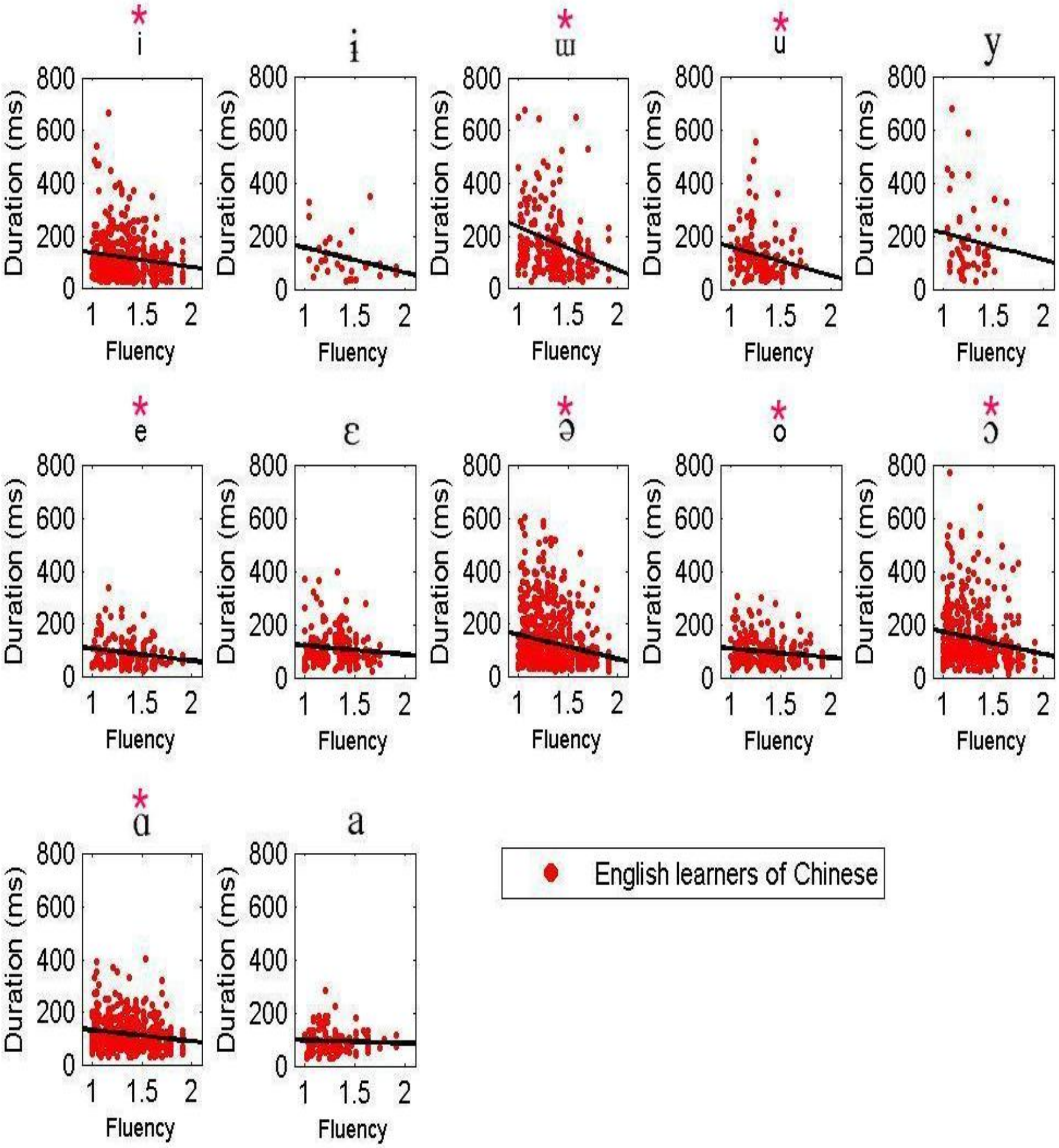


Figure 4.36: Correlation between vowel duration and fluency rating for English learners of Chinese. $*p < 0.05$.

Spectral Data

Due to the difference of formant frequencies between genders and the unbalanced distribution of gender in the corpus, the spectral data were separated by gender for analysis (native speakers: 7 females and 2 males; heritage learners: 5 females and 12 males; English learners: 5 females and 15 males). For each vowel, formant frequencies were extracted at the stable midpoint of the vowel duration and then averaged over all the snippets produced by the same speaker. The means and standard deviations of female productions are given in Table 4.19; those of male production are given in Table 4.20.

Figure 4.37 and Figure 4.38 illustrated the vowel space produced by speaker groups separated by gender, suggesting that Mandarin vowels posed different degrees of difficulty for the L2 learners. The mid vowels [e, ɛ, ə, ɔ, o] produced by heritage and English learners patterned closely to that by native speakers in both male and female productions. In female productions, the F2 of high vowels [i] and [y] produced by native speakers was between heritage and English learner production, while they are both more advanced (higher F2) in male native productions. According to the comparison of vowels between Mandarin and English in Chapter 3, the major difference between these two languages is that Mandarin vowels are further back than English vowels. Another observation is that Mandarin [u] is more rounded than English [u], which causes lower formant values. This vowel property can be observed in the [u] production by female L2 learners. They produced Mandarin [u] more fronted than female native speakers did, indicating that L2 learners carried L1 acoustic properties when they produced L2 sounds.

As for the low vowels, the comparison between English [ɑ] and Mandarin [a, ɑ] does not show significant formant differences in both F1 and F2, while [a] and

Vowel/F1	Native Mandarin	Heritage learners	English learners
/i/	414 (138)	433 (65)	396 (106)
/ɨ/	473 (195)	440 (62)	558 (72)
/ɯ/	449 (125)	462 (153)	475 (140)
/u/	464 (154)	427 (80)	418 (82)
/y/	384 (76)	406 (87)	443 (179)
/e/	527 (152)	510 (121)	515 (113)
/ɛ/	521 (165)	541 (144)	522 (150)
/ə/	549 (201)	524 (159)	518 (175)
/o/	535 (139)	542 (139)	526 (151)
/ɔ/	544 (165)	537 (132)	532 (133)
/a/	737 (227)	709 (153)	664 (183)
/ɑ/	790 (183)	710 (193)	719 (182)
Vowel/F2	Native Mandarin	Heritage learners	English learners
/i/	1924 (582)	1872 (501)	1989 (462)
/ɨ/	1745 (345)	1719 (270)	1686 (73)
/ɯ/	1850 (290)	1695 (311)	1783 (220)
/u/	1381 (396)	1474 (323)	1487 (329)
/y/	1876 (503)	1988 (449)	1650 (335)
/e/	2029 (421)	1968 (401)	2095 (305)
/ɛ/	2033 (423)	1921 (333)	1973 (234)
/ə/	1722 (321)	1710 (302)	1731 (277)
/o/	1359 (310)	1456 (351)	1355 (297)
/ɔ/	1391 (340)	1552 (350)	1441 (302)
/a/	1704 (293)	1591 (335)	1597(236)
/ɑ/	1610 (285)	1632 (339)	1656 (239)

Table 4.19: Mean female formant frequencies (in Hertz) for speaker groups of Mandarin native speakers, heritage speakers and English-speaking learners. The upper panel presents F1 values and the lower panel presents F2 values. Standard deviations are given in parentheses.

Vowel/F1	Native Mandarin	Heritage learners	English learners
/i/	358 (76)	361 (84)	348 (84)
/ɨ/	383 (64)	378 (75)	416 (91)
/ɯ/	398 (62)	402 (93)	397 (85)
/u/	414 (88)	368 (98)	352 (135)
/y/	485 (309)	342 (74)	334 (63)
/e/	421 (77)	432 (56)	448 (74)
/ɛ/	465 (80)	466 (120)	462 (107)
/ə/	459 (95)	441 (110)	462 (132)
/o/	459 (74)	462 (119)	489 (113)
/ɔ/	453 (77)	456 (112)	451 (132)
/a/	564 (134)	541 (125)	610 (181)
/ɑ/	600 (114)	580 (119)	639 (149)

Vowel/F2	Native Mandarin	Heritage learners	English learners
/i/	1942 (343)	1845 (307)	1818 (386)
/ɨ/	1564 (193)	1527 (110)	1506 (177)
/ɯ/	1504 (230)	1520 (211)	1539 (189)
/u/	1393 (466)	1381 (416)	1347 (433)
/y/	1872 (427)	1704 (269)	1613 (310)
/e/	1803 (299)	1743 (225)	1723 (200)
/ɛ/	1810 (217)	1692 (253)	1703 (190)
/ə/	1483 (281)	1504 (237)	1493 (256)
/o/	1206 (396)	1323 (352)	1188 (321)
/ɔ/	1211 (314)	1281 (355)	1224 (353)
/a/	1550 (251)	1411 (202)	1410(227)
/ɑ/	1377 (252)	1379 (221)	1370 (229)

Table 4.20: Mean male formant frequencies (in Hertz) for speaker groups of Mandarin native speakers, heritage speakers and English-speaking learners. The upper panel presents F1 values and lower panel presents F2 values. Standard deviations are given in parentheses.

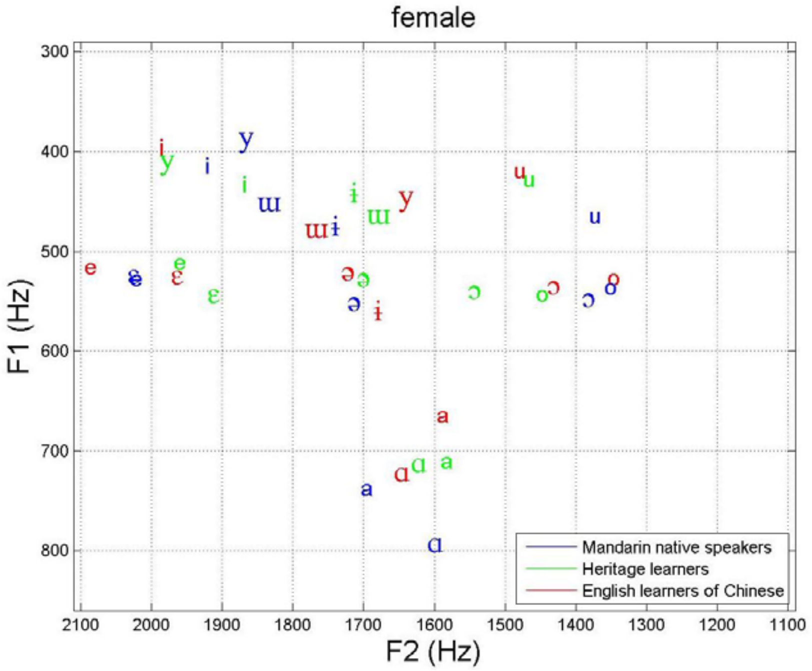


Figure 4.37: Vowel Space of the mean formant values by female speaker groups in the classroom data.

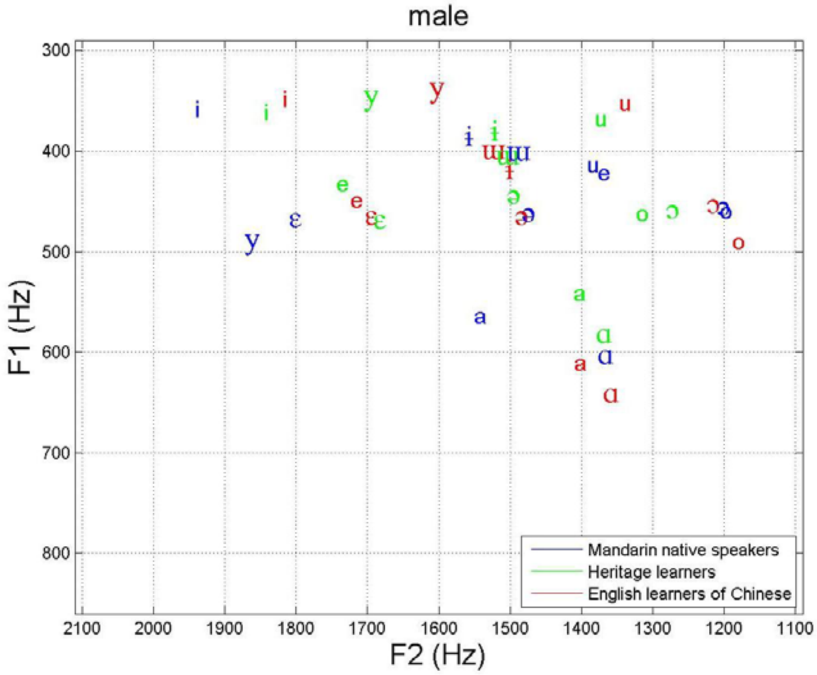


Figure 4.38: Vowel Space of the mean formant values by male speaker groups in the classroom data.

Speaker groups	Front (high F2)	Back (low F2)	High (low F1)	Low (high F1)
Female native speakers	a	ɑ	a	ɑ
Female heritage learners		a, ɑ		a, ɑ
Fenake English learners	ɑ	a	a	ɑ
Male native speakers	a	ɑ	a	ɑ
Male heritage learners		a, ɑ	a	ɑ
Male English learners		a, ɑ	a	ɑ

Table 4.21: Summaries of the low vowels production by speaker groups

[ɑ] in Mandarin native production do have significant differences in F2 ([a] is more fronted than [ɑ] because of the coarticulation effect). Similarly, in the spontaneous speech, [a] has a lower F1 and a higher F2, indicating that [a] is higher and more fronted than [ɑ] in both female and male native production. Interestingly, female English learners produced these two vowels in the opposite way that native speakers did, suggesting that they are aware of the distinction between these vowels but switched the vowel space. In female heritage productions, they did not distinguish these two vowels in terms of formant values. For both male heritage speakers and male English learners, it seemed that they were able to distinguish these two vowels in F1 (vowel height), but not in F2 (vowel backedness). Both [a] and [ɑ] in the productions of male heritage speakers and male English learners have similar F2 values. This implies that L2 male speakers use the wrong dimension to distinguish these two vowels in Mandarin. They might produce both of the Mandarin low vowels more like English [ɑ]. In addition, the vowel space of [a] and [ɑ] in male English learner productions is much lower (higher F1) than for native speaker and heritage learners. The observations of low vowels are summarized in Table 4.21.

The SLM predicted that similar, but not identical phones between L1 and L2 should be difficult for L2 learners to acquire because of the effect of equivalence classification, while sounds different from L1 categories should be easier to be learned eventually. Based on vowel similarity between Mandarin and English presented in Chapter 3, Mandarin vowels were classified into four categories.

- Vowels [i, e, ɛ, o, a ɑ] are the same vowels in Mandarin and English (they will not cause problem for L2 learners).
- Mandarin [u, ə] are similar, but not identical to English counterparts (they are the most difficult ones).
- Mandarin [ɔ] is a different vowel compared to English counterpart (it is relatively easy to establish an L2 category for this sound).
- Mandarin [y, ɨ, ɯ] are new vowels for L2 learners (it is comparatively easy to build new categories for these vowels).

In the vowel productions of the corpus data, L2 learners indeed do not have trouble with Mandarin [i, e, ɛ, o], as the SLM predicted, while it is very difficult to learn the Mandarin low vowels [a] and [ɑ]. One possible reason might be the mapping problem of two L2 sounds to one L1 sound, as the female heritage learners did. Alternatively, learners are aware that there are two sounds in the L2, but they use the wrong dimension to distinguish them or map them into two L1 categories ([æ] due to orthographic confusion and the English low vowel [ɑ]), as the female English learner. According to the SLM's predictions, another difficult sound is Mandarin [u] and it does cause difficulty in L2 learning, while the [ə] is not an issue for L2 learners. The L2 production of Mandarin [u] carried the L1 English color of [u], meaning that the tongue position in the L2 production of [u] was not as back as the [u] for native speakers nor were the lips rounded sufficiently. Mandarin [ɔ] is different from its English counterpart and it is easy for L2 learners to acquire, as shown in the data. As for another supposedly easy-learning group [y, ɨ, ɯ], they are new sounds for L2 learners. The [y] in L2 female productions shows higher F1 and lower F2, suggesting that [y] is neither rounded and nor fronted enough. The vowel [ɨ] in both male and female English learner productions is more close to to the mid vowel [ə]. The [ɯ] in both female and male learner productions is close to the [ɯ] in the native vowel space. The difficulty in learning [ɨ] and [ɯ] might result from the empty or unspecified category was reinforced by Pinyin. Another possible explanation for the difficulty in learning [ɨ, ɯ] might be due to the articulatory properties of these two sounds. The articulation of these two vowels carries over the tongue position of the preceding consonants, indicating that there is no open-close cycle for the CV sequence, such as [ta]. Thus, language learners have to learn not to move their tongue when producing these two vowels.

4.7 Summary

In this chapter, a questionnaire of 8 questions is used to evaluate second language spontaneous speech production. The results of both the rating experiment and the acoustic analysis shed light on our understanding of second language fluency and foreign accent. The data from the rating experiment revealed the mappings between perceptual judgements and acoustic measures through PCA. The vowel analysis showed a number of properties in the accented speech produced by L2 learners, which could be attributed to characteristics of the English vowel system.

The findings of the perceptual ratings indicated that the rating variables were all correlated in both the classroom and the picture telling data, although the fluency rating in the simplest task, clock telling, reached a ceiling effect. Other task types including simple picture telling, complex picture telling, and classroom data showed a consistent pattern among those rating scores despite the length of the snippets which ranged from 7 to 15 seconds. This suggests that native listeners are able to obtain a broad impression of learners' speech in a very short amount of time. The correlation analysis also demonstrated that disfluencies and accentedness are highly correlated with the ratings of vocabulary and pronunciation, respectively. The high correlations between disfluencies and vocabulary sizes in heritage and English learner groups suggests that disfluencies in L2 speech might relate to the lack of vocabulary. Through PCA, rating variables and acoustic measures were grouped into two categories to indicate fluency and accent by the second principal component (y axis). F1 and F2 of vowels were in the group of accentedness, nativeness, pronunciation. The other group contains variables including fluency, disfluency, vocabulary, grammar, and comprehensibility, with acoustic measures of AR, RS, PTR, FP number and FP duration. Among the acoustic measures, the RS and the F2 of vowels were the most powerful predictors investigated in the perception of fluency and accentedness, respectively. This indicated that F2 and RS contributed the most to the evaluation of fluency and foreign accent. Note that we are not interpreting F2 and RS as the only predictors of accentedness and fluency that native speakers used. There are other acoustic attributes in the prosodic domain that have not been investigated in this study. The linear regression analysis between vowel duration and fluency scores suggested that a slow speech rate caused lower fluency ratings in the production of English learners.

The vowel analysis of the corpus data revealed that F2 was a significant dimension for accentedness because L2 productions of [y, u, a, ɑ] differed considerably from native-speaker norms in F2. The production of [u] by learners demonstrated the effect of L1 transfer in L2 speech. The difficulty in learning Mandarin [a] and [ɑ] results from multiple factors, such as the phonological distribution in Mandarin, the coarticulation effect and Pinyin confusion. The new sound [y, ɨ, ɯ] poses difficulty in different degrees. The [y] production is not fronted and high enough in L2 speech. The production of [ɨ] is close to [ə]. The results from the analysis of the spontaneous speech failed to entirely support SLM's prediction. The vowel analysis shows that segmental similarity is not the only factor to predict L2 sound productions. Other factors should be taken into account as well.

Chapter 5

General Discussion

In this thesis, a questionnaire of eight questions was used to evaluate spontaneous speech productions of second language speech. The results of both the human-rating experiment and the acoustic analysis shed light on our understanding of second language fluency and foreign accent. The rating data revealed the mappings between perceptual judgements and acoustic measures through PCA. The vowel analysis demonstrated a number of properties in the accented speech produced by L2 learners, which could be attributed to characteristics of English vowel system.

5.1 Fluency

In the perceptual rating experiment, the correlation analysis revealed that the disfluency rating is highly correlated with the vocabulary rating, especially in L2 production. The PCA analysis provided an index formulated by the original rating variables to evaluate how good the speech is. The PC1 is composed of all of the rating variables with a similar positive weight. The PC2 shows clean separation of rating questions into two groups. One of the groups, consisting of the variables fluency, disfluency, vocabulary, grammar, and comprehensibility, is more relevant to the knowledge of the language, which affects the formation of the intended message. The acoustic measures, loaded on the platform of the PCA-rating results, illustrated that the rate of speech, the articulation rate, the phonation time ratio, the standard deviation of the vowel duration, and the percentage of vowel duration were classified with the knowledge factors. Among the acoustic attributes, rate of speech is the best predictor of fluency.

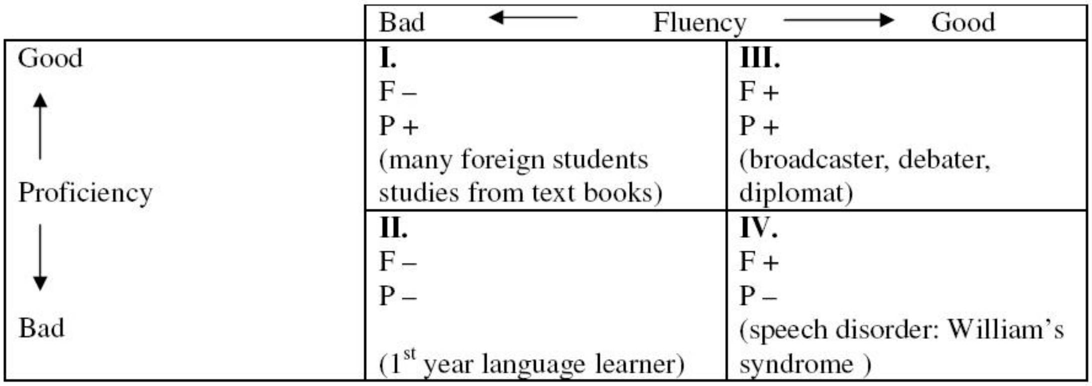
What is fluency? What definition and mechanism of fluency can be provided from the current study? The literature started off by defining fluency as flowingness. Disfluencies disrupt this flow. However, disfluencies also occur in native speech, however they do not impede communication. Thus, disfluency is not opposite of fluency. Most studies agree that the rate of speech is correlated to fluency. Thus, what attributes in the acoustic speech cause listeners to perceive an utterance as fluent? What attributes of speakers cause speakers to be able to produce fluent speech? What aspects of fluency are specific to L2 learners and what aspects are specific to native speakers?

The acoustic attributes, such as speaking rate, frequency of filled pauses, ect., are the surface characteristics of fluency, which exist in both native and non-native speech. The findings suggest that these attributes causes listeners to determine whether speech is fluent or not. On the other hand, what allows speakers to produce fluent speech with high speaking rates? The mechanism of fluency of native speakers and learners might be entirely different.

In native productions, the correlation between disfluency ratings and word types (word types = 33) was lower at 0.41 than for the heritage speakers and English speakers. In addition, there was no significant correlation between word counts (word counts = 47) and disfluency rating. Obviously, vocabulary size is not a crucial attribute of L1 fluency. Then, what causes native fluency? Studies related to the delayed auditory feedback effect demonstrated that disfluencies are induced and the speech rate decreased in normal fluent native speech when speech production is not synchronized with its feedback to the auditory system (Fabbro & Darro, 1995; Van Borsel, Sunaert, & Engelen, 2005; Chon, 2010). Chon(2010) showed that the articulation rate (from 5.10 syllable per second to 3.11 syllable per second) of fluent native speech dropped significantly during the disruption of delayed auditory feedback. Interestingly, the articulation rate fell into the range as the articulation rate (3.28 syllables per second) of English learners reported in the current study. Although the mechanisms explaining the link between auditory feedback and fluent speech production is unclear, the Directions into Velocities of Articulators (DIVA) model provides a theory to explain that the time lags of auditory feedback control can causes disfluencies in speech movement (Perkell et al., 2000; Guenther & Perkell, 2004). Thus, in native speech, delay or distortion of the auditory feedback loop causes difficulty in producing fluent speech. Because delayed auditory feedback effects are not influenced by vocabulary size, speech rate is the major attribute contributing to the perception of native fluency.

As for L2 fluency, the results of this study imply that vocabulary size is a big factor to language learners. The correlation analysis showed that disfluency rating highly correlated with vocabulary rating. Based on the vocabulary size used in 15-seconds of speech, English learners had comparatively small word types (17.12, $r = 0.76$) and word counts (25, $r = 0.73$), which significantly correlated with the disfluency rating. Similarly, the vocabulary size of heritage learners statistically correlated with the disfluency rating (word types = 25, $r = 0.72$; word counts = 35, $r = 0.64$). In speech planning, vocabulary is the building blocks of sentences. If a learner has difficulty in retrieving words, disfluencies may occur. In addition, learning a second language for adult learners may differ from babies learning to speak. A trait of adult learner speech is that they have the “thoughts” in L1 and that they then translate the meaning or words into the L2. This might slow down the speaking rate as well as cause disfluencies when they are searching for words. If a learner has a large enough vocabulary to respond to communicative needs, a good command of grammar to connect phrases and sentences, and an accurate pronunciation, will he/she be judged as a fluent speaker? L2 fluency has a lot to do with knowledge and the processing of L2 speech.

Another possibility for the high correlations of the rating variables might be due to some of variables correlating with proficiency–as a result, they correlate with each other. The knowledge-related group of the variables (fluency, disfluency, vocabulary, grammar, comprehensibility) might be related to the learner's proficiency. Proficiency is a composite measure of learners' abilities including competence (over the lexicon, grammar, and discourse) and performance (spoken and written) (Tremblay & Garrison, 2010; Tremblay, in press). It is also an approximation of the learners' abilities, whereas proficiency focuses on the learner's production. In this study, the rating experiment emphasized the role of the listener's perception and provided judgments of the learners' oral proficiency. In addition, acoustic measures quantitatively examined the temporal properties of learners' speaking performance. The calculated word types and word counts in speech production are an estimation of learners' proficiency of lexical competence. In the learner's lexical competence, raters respond to the appropriateness of lexical usage. There are numerous methods to estimate learners' abilities. One approach is to standardize the tests, through the use of cloze tests, such as in the TOEFL and ACTFL oral proficiency tests. Another is to randomly select speech samples from the learners' speech performance, which are representative and generalizable to the same speakers. This study adopted the later method to represent the learner’s proficiency of speech production and lexical competence, through the use of perceptual ratings.



	Bad ← Fluency → Good	
Good ↑ Proficiency ↓ Bad	**I.** F – P + (many foreign students studies from text books)	**III.** F + P + (broadcaster, debater, diplomat)
	II. F – P – (1st year language learner)	**IV.** F + P – (speech disorder: William’s syndrome)

Figure 5.1: The relationship between fluency and proficiency

The relationship between fluency and proficiency can be outlined as shown in Figure 5.1. The first category contains the speakers who have a good proficiency level, but who cannot produce fluent speech, e.g. many foreign students who study the language from text books for many years. If they are tested in a written proficiency test, they are probably able to obtain pretty good scores for vocabulary and grammar, while it is difficult for them to speak fluently. The second category contains speakers with low fluency and low proficiency. The first year language learners may belong to this category because they are still in the initial stage of learning a language. Thus, their proficiency level, in terms of lexical competence, grammatical competence, and speaking ability, is low. The third category contains the speakers with good fluency and good proficiency, e.g. broadcasters, debaters, and diplomatic ambassadors. They usually have excellent command of speaking and knowledge of the language. The fourth category is the speakers who are able to speak fluently, but the content of the intended message is nonsense. For example, most individuals with Williams syndrome have strong language skills. If strings of words are produced with a good flow, will this kind of speech be judged as fluent? The discussion here provides ideas for future studies to compare L1 and L2 fluency in a wider range with more extreme cases.

In sum, fluency is not simply related to speaking rates or to lexical competence. The definition of fluency should take into account both the listener's and the speaker's sides. Furthermore, the speech planning of fluency in native and non-native speakers might be caused by different factors.

5.2 Foreign Accent

In the correlation analysis, ratings of nativeness, accentedness, and pronunciation had a strong correlation with one another. The PCA analysis also confirmed that these three variables belong to the same group. The discrepancy of accentedness ratings for native speakers supported the notion of accent, which is defined as the deviation from speaker's norm. Thus, the Mandarin speakers from Taiwan gave lower rankings towards the speech produced by speakers from mainland China than to other Mandarin speakers from Taiwan. We can expect that the result would be opposite if the raters were speakers from mainland China. The relationship between accentedness and pronunciation suggested that it was easier to improve pronunciation, but more difficult to change the impression of accentedness. Learners were able to obtain high pronunciation scores, but the ratings of accentedness was still low.

From the vowel comparison between English and Mandarin in Chapter 3, the primary difference between vowels in Mandarin and English is in F2. The

PCA result also revealed that F2 is closer to the group of the factors related to oral performance. However, a crucial issue in the SLA theories is: how are the crosslinguistic similarities defined? The methods of assessing segmental similarities include the perceptual discrimination task, spectral analysis and the statistic training model. The SLM mainly uses phonetic properties to compare segmental similarities, while the PAM uses the perceptual task to infer the influence of articulatory gestures on perception. The PAM model does not explain which articulatory properties are easier or harder to learn. On top of these approaches, the articulaory properties are seldom used probably due to the difficulty of collecting articulatory data. Neither the SLM nor the PAM address the following problems:

- Which attributes should be used to define similarities? Perception? Acoustics? Articulation? If there are discrepancies in the comparison of the similarities across languages, which one should be the criterion?
- In the analysis, which values should be used to assess the similarities? mean? median? the population in the scatter plot?
- Beyond the comparison of segmental similarities, what other factors, such as context differences, should be take into account to explore learners' behaviour? For example, Mandarin has context variations, which influence the phonetic qualities of vowels to a certain degree.

The Mandarin vowels [i, e, o, ɛ, ɑ, a] can be regarded as either spectrally identical or similar to their English counterparts, depending on whether the speaker that produced the utterances is male or female. If they are identical, one might expect that these vowels were easy to learn in Mandarin based on the predictions of SLM. If they are similar, then, one might expect that these vowels would be difficult to learn. Thus, the hypotheses of SLM is difficult to test.

In Mandarin, [y] is a high front rounded vowel, which is a new sound for L2 learners. [u] is a high back rounded vowel but it is further back than the English

[u]. The L2 production of [y] is lower and further back than the native production and is closer to their L2 production of [u]. [u] in L2 productions is more fronted (higher F2) and less rounded (rounding will lower formant values globally) than native productions, which is similar to English [u]. These observations reveal the L1 transfer effect on L2 speech. Moreover, the constraint between [+back] and [+rounded] is strong in English and leads to [u]-like production of [y] in L2 speech. L2 learners struggle to disassociate the articulatory constraint that violates the articulatory pattern in English.

Mandarin vowels [ɨ] and [ɯ] only occur, respectively, with alveolar sibilants and post-alveolar retroflex, which are new vowels for L2 learners. The results showed that the production of [ɨ] by English learners is closer to Mandarin [ə]. The articulatory properties of [ɨ] and [ɯ] maintain the tongue position of the preceding consonants, which violate the articulation of CV open-close oscillation. Thus, a L2 learner needs to learn not to move his/her tongue when producing these two vowels.

The low vowels [a, ɑ] in Mandarin create great difficulty for language learners. The articulatory finding described that articulation of these low vowels is influenced considerably by the surrounding consonants. Phonetically, due to coarticulation effect, these two sounds have different vowel qualities, particularly in F2 values. However, they are indicated with the same letter *a* in the transliteration system, Pinyin, which induces English learners to pronounce Mandarin low vowels as English [æ]. The comparison of vowels in these two languages found that the English low vowel [ɑ] is identical to the Mandarin [a] or [ɑ]. All of these factors make Mandarin low vowels difficult to learn. In the findings, female heritage learners produced both Mandarin low vowels more like the vowel [ɑ], indicating that they did not learn the differentiation of these two sounds. Male L2 learners also produced Mandarin low vowels as back as [ɑ] with slightly F1 differences (vowel height), implying that they might learn the discrepancies between these two phones, but that they use the wrong dimension to distinguish them. Female L2 learners flipped the direction of Mandarin low vowels, suggesting that they learned the contrast of these two vowels, but switched the categories.

Consequently, the findings failed to support the predictions of SLM because the L2 learners did succeed in learning similar vowels and had problems in learning new vowels, as well as identical vowels. The behaviour of L2 vowel production is beyond the similarity measure of vowels. Other factors, such as articulatory properties, coarticulation effect, transcription confusion, should be taken into account for L2 pronunciation. The findings show that there are different stages in the acquisition of Mandarin vowels and how L1 pronunciation contributes to the perception of a foreign accent.

Chapter 6

Conclusion

6.1 Summary of the Findings

This study investigated the factors contributing to the perception of second language fluency and foreign accent, combining an acoustic analysis of conversational speech samples and native listeners' rating scores. About 400 speech samples were randomly selected from speech corpora, where speakers performed a variety of tasks in natural speech. Forty-three linguistic naïve raters listened to each speech sample and answered 8 questions about their impression of the speech that they heard. They first reported their broad impressions as to whether the speech sounded fluent, whether the speaker had an accent, and whether the speaker was a native speaker of Chinese. Next they evaluated specific aspects of the speech sample: how appropriate was the vocabulary choice, how accurate was the grammar, how good was the pronunciation, and whether the speech was easily comprehensible. A correlation analysis, a PCA and the analysis of acoustic measures were performed to address the research questions and to investigate what might influence native listeners' perception of fluency and foreign accent and what causes speakers to produce fluent speech. Below is a summary of the findings corresponding to the research questions.

1. What leads to the perception of second language fluency and foreign accent?

Because of the high correlation among rating variables, PCA was used for dimension reduction. The results showed that all rating factors contribute similarly to the PC1. The PC2 classified the rating variables into two groups. One group consisted of the variables fluency, disfluency, grammar, vocabulary, and comprehensibility, representing the knowledge factors that construct the speech. The other group consisted of the variables of nativeness, accentedness, and pronunciation, representing sound-related factors that contribute to the perception of foreign accent.

A detailed examination of individual scoring patterns demonstrated that disfluencies are well correlated with vocabulary. Speakers with a small vocabulary size, including some of heritage learners and nearly all of the English speaking learners, have more severe disfluencies, implying that difficulty in lexical access might lead to disfluency in L2 speech, which in turn leads to a low fluency rating. Improvements in grammatical knowledge and pronunciation do not necessarily lead to improvement in fluency. This result is consistent with a study by Towell and Dewaele (Towell & Dewaele, 2005). Some L2 speech was ranked low in fluency, while ranking high in grammar or pronunciation. At the same time, some L2 learners received a high fluency score, but received a low score in pronunciation or grammar.

Accent is the perceptual distance between the speech of the listeners and the speakers. The native listeners from Taiwan recognized a dialectal accent in the speech produced by native speakers from Mainland China. Hence, raters ranked the dialectal accent lower due to the difference from their own speech. If the raters were from China, we would expect the opposite scoring patterns. Raters scored accent rather harshly. We expected native speakers to receive high scores on all questions. This is true except for their accent scores. Many native speakers received low scores on accent, and all of them were from mainland China, while the raters were from Taiwan.

The relationship between accent and pronunciation was also examined. It was found that it is easier to improve the pronunciation, while the impression of accent is hard to change. In general, accent scores are lower than pronunciation scores.

> 2. What is the relationship between acoustic measures and human-rated judgements of fluency and foreign accent?

Rate of speech is the best predictor of fluency, while other acoustic attributes, such as the articulation rate, the phonation time ratio, and the number and duration of filled pauses, are also good predictors. The rate of speech and the vowel duration both reflect that the faster the speaking rate, the higher the fluency rating.

The F2 of vowels is a strong predictor of foreign accent. The major difference between Mandarin and English vowel systems lies in the backedness of the vowel quality (F2 dimension). Some of the L2 vowel productions carry the color of their L1 acoustic characteristics.

> 3. How to measure vowel similarities? How similar are Mandarin and English vowels?

There is no consensus in the literature about the phonological vowel inventory of Mandarin. In this thesis, the largest number of phonetic vowel categories were used in assessing the similarity between Mandarin and English vowel inventories and their production by different speaker groups. Phonetically, there are five high vowels [i, ɨ, ɯ y, u], five mid vowels [e, ɛ, ə, o. ɔ] and two low vowels [a, ɑ] in Mandarin. The articulatory study of Mandarin vowels demonstrated that the tongue body positions were different for the vowels [i, ɨ, ɯ], while formant values of F2 depicted that [ɨ, ɯ] were significantly different from [i]. There was no difference between [ɨ] and [ɯ].

All phonological studies of Mandarin assume one low vowel. We found that the two low vowels [a] and [ɑ] have different articulatory properties due to the coarticulation effect. L2 learners encountered a lot of problems in the production of [a, ɑ]. This is an area where targeted training is needed.

The comparison of vowels between Mandarin and English revealed that the major differences lay in the fact that Mandarin vowels are further back than English vowels. In addition, it was observed that Mandarin [i], [e], [o], and [ɛ] are identical to their English counterparts. Mandarin [u] is articulated further back and is rounded more than English [u]. Mandarin [ɔ] is very different from English [ɔ]. Vowels [y, ɨ, ɯ] do not exist in English and thus are new vowels for the L2 learners. The SLM (Flege, 1995b) predicts that similar, but not identical vowels are difficult to learn, while it is easier to establish new categories for different or new vowels. Identical vowels should not cause problems in L2 pronunciation acquisition. Hence, it should be easier to learn Mandarin vowels [y, ɨ, ɯ, ɔ] and [i, o, e, ɛ, a, ɑ] and difficult to learn Mandarin [u].

> 4. How do L2 learners produce vowels in Mandarin? Do Mandarin vowels pose different levels of difficulty to L2 learners? What are the factors that influence L2 vowel production?

The vowel space of Mandarin native speakers, heritage speakers and English speaking learners were different. The formant space of vowels [i, o, e, ɛ, ə, ɔ] for L2 learners appeared to pattern closely with the productions of Mandarin native speakers. The difficult Mandarin vowel [u] in L2 production showed more advanced vowel positions (higher F2) than that of the L1 productions, suggesting that this vowel was influenced by properties of the L1 (English).

In contradiction to the SLM's predictions, Mandarin [a] and [ɑ] were difficult for L2 learners. The formant space of [a] and [ɑ] produced by female English learners were in the opposite direction from the way Mandarin female native speakers produced them. Female heritage learners did not distinguish these two vowels in their productions. In male L2 productions, both vowels were pronounced as equally back as native productions, but they distinguished these two vowels slightly in their F1. All of these phenomena imply that some of the language learners use the wrong dimension to distinguish Mandarin low vowels. Others mapped these two vowels into one categories as in their L1. In Pinyin, both low vowels [a, ɑ] are written with the letter *a*, which does not specify a coarticulation effect in the vowel quality. Furthermore, the vowels in the spelling of *an* and *ang* (e.g. tan, tang) are all pronounced as [æ] in English, while they are pronounced differently in Mandarin ([a] in 'tan' and [ɑ] in 'tang'. The findings revealed that the L2 vowel production is affected by the L1 sound system, allophonic variations in the L2, and transliteration confusion.

Mandarin [y], a front rounded vowel, is difficult for L2 learners who automatically associate front vowels with unrounding. Some L2 learners produced [y] with rounding, but they retracted their tongue, resulting in a sound that falls between [y] and [u]. Additionally, they did not produce the vowel high enough.

Mandarin [ɨ] in L2 productions was closer to [ə], indicating that learners had trouble in pronouncing [ɨ] in Mandarin. In the transliteration system, Pinyin, the three high vowels [i, ɨ, ɯ] are indicated with the same letter *i*. Learners had difficulty in differentiating the sounds, which is erroneously reinforced by the transcription system. This accounts for the difficulty in acquisition. SLM predicts that a new sound is easy to learn. This prediction is not born out, perhaps because of the new coordinate of articulatory movement.

6.2 Implications of the Current Study

One of the most important differences in this study, compared to previous studies, is that the findings are based on continuous spontaneous speech data. Spontaneous speech is closer to the natural speech that people speak on a daily basis. In addition, this study examined these two topics by integrating acoustic measures and human-rated perceptual judgements. The results bridge the gap and work out a detailed mapping relationship between speech production and speech perception. Comparing vowel similarities between Mandarin and English will allow us to identify crosslinguistic acoustic properties through which we can examine predictions from SLA theories.

A better understanding of second language fluency and foreign accent will eventually allows us to design a corrective method to reduce accent and improve fluency. This work aids second language instructors by identifying the difficult areas of Mandarin vowel productions by L2 speakers and raises awareness about reasons for sound confusion. Pronunciation training methods can be designed to target particular vowels.

In the field of language testing, understanding the acoustic attributes of fluency will help to improve the accuracy of automatic assessment systems that classify native and non-native speech or score second language fluency. Our study shows that evaluation of fluency and accent can be achieved with short speech samples.

6.3 Further Research

Fluency

This dissertation studies fluency in native as well as learners' productions. It is an interesting question to extend the research to study the differences between L1 and L2 fluency. Disfluencies exist in native speech and they usually do not impede communication. What are the differences between L1 disfluency and L2 disfluency? How is L2 disfluency different from L1 fluency disorders? How can we quantify the aspects that influence the perception of fluency? Is it slower speaking rate or perhaps the number, duration, location, or type of hesitation, or silent or filled pauses that distinguish them? In this study, native speakers reached the ceiling and the English learners were at the floor in the rating scale. The native speakers and learners in the corpus data do not cover the complete range of fluency nor the other metrics observed. One could conduct an experiment, including speech samples from native and learners' speech covering extremely fluent to extremely disfluent speech and measure the acoustic attributes to investigate the relationship between them.

Accent

We found that listeners detected dialectal accents between Taiwanese Mandarin and Mainland Mandarin even with a small number of native speakers in this study. Future studies could enlarge the pool of speakers and raters to further investigate this phenomenon. Other factors, such as tone acquisition and intonation could be examined to study foreign accent in L2 production.

Vowel Studies

The vowel study in Section 3 allows us to explore the acoustic and articulatory properties of vowels in Mandarin, but it is still not clear how native Mandarin listeners perceive the vowel categories. Further studies could design a perceptual task to examine native speaker's perceptual map of Mandarin vowel categories. Then, an integrated study could compare the similarities and discrepancies of Mandarin vowel categories in terms of phonetics, phonology and articulatory positions.

I would also like to extend this articulatory study to English vowels produced by English native speakers to compare the articulatroy positions of English and Mandarin vowels. In the literature, vowel similarities were assessed based on the perceptual discrimination task, spectral comparison and statistical pattern recognition models. Evaluation of vowel similarities by comparing the articulatory differences would contribute to the discussion.

The articulatory data of Mandarin vowels produced by English learners allow us to compare the kinematic pattern in native and non-native speech that may be the underlying cause of L2 production difficulties.

Combing the articulatory and acoustic data from native production of Mandarin and English and L2 production of Mandarin vowels, this allows us to further compare the SLM and PAM models and examine the behaviour of L2 speech.

References

AG500 manual. (2006). Medixinelektronik GmbH.

Ambady, N., & Rosenthal, R. (1993). Half a minute: Predicting teacher evaluations from thin slices of nonverbal behavior and physical attractiveness. *Journal of Personality and Social Psychology*, *64*(3), 431-441.

Anderson-Hsieh, J., Johnson, R., & Koehler, K. (1992). The relationship between native speaker judgments of nonnative pronunciation and deviance in segmentals, prosody, and syllable structure. *Language Learning*, *42*(4), 529-555.

Au, T. K.-f., Knightly, L. M., Jun, S.-A., & Oh, J. S. (2002). Overhearing a language during childhood. *Psychological Science*, *13*, 238-243.

Best, C. T. (1995). *A direct realist view of cross-language speech perception.* In W. Strange (Ed.), *Speech perception and linguistic experience: Theoretical and methodological issues* (p. 121-154). Timonium, MD: York Press.

Best, C. T., McRoberts, G. W., & Goodell, E. (2001). Discrimination of non-native consonant contrasts varying in perceptual assimilation to the listener's native phonological system. *The Journal of the Acoustical Society of America*, *109*(2), 775.

Bhat, S. P. (2010). *Estimation problems in speech and natural langauge.* Unpublished doctoral dissertation, University of Illinois at Urbana-Champaign.

Birdsong, D. (2005). Interpreting age effects in 2nd language acquisition. In J. Kroll & A. d. Groot (Eds.), *Handbook of bilingualism.* (p. 109-127). New York: Oxford University Press.

Bock, J. K., Brehm, L., Kuchinsky, S., Lam, T., Lee, C., Ospina, J., et al. (2010). *Putting words together into simple phrases, uh, fluently.* University of Illinois, Urbana-Champaign, Nov 5, 2010.

Bock, J. K., Irwin, D. E., Davidson, D. J., & Levelt, W. J. M. (2003). Minding the clock. *Journal of Memory and Language*, *48*, 653-685.

Bohn, O.-S., & Flege, J. E. (1992). The production of new and similar vowels by adult German learners of English. *Studies in Second Language Acquisition*, *14*(2), 131-158.

Bradlow, R., Ann, & Bent, T. (2008). Perceptual adaptation to non-native speech. *Cognition*, *106*, 707-729.

Breiman, L., Friedman, H., Olshen, R. A., & Stone, S. J. (1984). *Classification and regression tree.* Belmonk, CA: Wadsworth.

Browman, C., & Goldstein, L. M. (1986). Toward an articulatory phonology. *Phonology Yearbook*, *3*, 219-252.

Buck, K. (1999). *ACTFL oral proficiency interview tester training manual.* ACTFL, Inc.

Chafe, W. (1980). *The paper stories: Cognitive, cultural and linguistic aspects of narrative production.* Norwood, New Jersey: Ablex.

Chao, Y. R. (1968). *A grammar of spoken Chinese.* Berkeley, CA: University of California Press.

Chen, K.-J., & Bai, M.-H. (1998). Unknown word segmentation for Chinese by a corpus-based learning method. *International Journal of Computational linguistics and Chinese Language Processing*, *3*(1), 27-44.

Cheng, C.-C. (1968). English stresses and Chinese tones in Chinese sentences. *Phonetica*, *18*, 77-88.

Cheng, C.-C. (1973). *A synchronic phonology of Mandarin Chinese synchronic phonology of Mandarin Chinese.* The Hague: Mouton.

Chomsky, N., & Halle, M. (1968). *The sound pattern of English.* New York: Harper Row.

Chon, H.-C. (2010). *Auditory-motor integration influences on speech motor control and fluency: A comparison of normally fluent speakers and people who stutter*. Unpublished doctoral dissertation, University of Illinois at Urbana-Champaign.

Clark, H. H., & Fox Tree, J. E. (2002). Using uh and um in spontaneous speaking. *Cognition*, *84*(1), 73-111.

Clements, N. G. (1985). The geometry of phonologial features. *Phonology Yearbook*, *2*, 225-252.

Cucchiarini, C., Strik, H., & Boves, L. (2000). Quantitative assessment of second language learners' fluency by means of automatic speech recognition techonology. *Journal of Acoustical Society of America*, *107*(2), 989-999.

Cucchiarini, C., Strik, H., & Boves, L. (2002). Quantitative assessment of second language learner's fluency: Comparing between read and spontaneous speech. *Journal of Acoustical Society of America*, *111*(6), 2862-2873.

Derwing, T. M., Munro, M. J., Thomson, R. I., & Rossiter, M. J. (2009). The relationship between L1 fluency and L2 fluency development. *Studies in Second Language Acquisition*, *31*(04), 533-557.

Derwing, T. M., Thomson, R. I., & Munro, M. J. (2006). English pronunciation and fluency development in Mandarin and Slavic speakers. *System*, *34*(2), 183-193.

Duanmu, S. (2000). *The phonology of standard Chinese.* Oxford New York: Oxford University Press.

Fabbro, F., & Darro, V. (1995). Delayed auditory feedback in polyglot simultaneous interpreters. *Brain and Language*, *48*, 309-319.

Fant, G. (1960). *Acoustic theory of speech production.* Mouton: The Hague.

Flege, J. E. (1988). Factors affecting degree of perceived foreign accent in English sentences. *Journal of Acoustical Society of America*, *84*(1), 70-79.

Flege, J. E. (1995a). Factors affecting strength of perceived foreign accent in a second language. *Journal of Acoustical Society of America*, *97*(5), 3125-3134.

Flege, J. E. (1995b). *Second language speech learning: Theory, findings and problems.* In W. Strange (Ed.), *Speech perception and linguistic experience: Theoretical and methodological issues* (p. 121-154). Timonium, MD: York Press. York Press.

Flege, J. E. (2003). Interaction between the native and second language phonetic subsystems. *Speech Communication*, *40*(4), 467-491.

Flege, J. E., MacKay, I. R. A., & Piske, T. (2002). Assessing bilingual dominance. *Applied Psycholinguistics*, *23*, 567-598.

Goffiman, E. (1981). *Radio talk.* Philadelphia, PA: University of Pennsylvania Press.

Griffin, Z. M., & Bock, J. K. (2000). What the eyes say about speaking. *Psychological Science*, *11*, 274-279.

Guenther, F. H., & Perkell, J. S. (2004). *A neural model of speech production and its application to studies of the role of auditory feedback in speech.* Oxford, United Kingdom: Oxford University Press.

Guion, S. G., Flege, J. E., Akahane-Yamada, R., & Pruitt, J. C. (2000). An investigation of current models of second language speech perception: The case of Japanese adult's perception of English consonants. *Journal of Acoustical Society of America*, *107*(5), 2711-2724.

Hammerly, H. (1982). Contrastive phonology and error analysis. *International Review of Applied Linguistics in Language Teaching*, *20*, 17-32.

Hasegawa-Johnson, M. A., Pizza, S., Alwan, A., Cha, J. S., & Haker, K. (2003). Vowel category dependence of the relationship between palate height, tongue height, and oral area. *Journal of Speech, Language, and Hearing Research*, *46*, 738-753.

Hellwig, B. (n.d.). *Elan-linguistic annotator. http://www.lat-mpi.eu/tools/elan/.*

Hillenbrand, J., Getty, L. A., Clark, M. J., & Wheeler, K. (1995). Acoustic characteristics of American English vowels. *Journal of Acoustical Society of America*, *97*(5), 3099-3111.

Hosoda, M., & Stone-Romero, E. (2010). The effects of foreign accents on employment-related decisions. *Journal of Managerial Psychology*, *25*(2), 113-132.

Howie, J. M. (1970). *The vowels and tones of Mandarin Chinese: Acoustical measurements and experiments.* Unpublished doctoral dissertation, Indianan University.

Hu, G., & Lindemann, S. (2009). Stereotypes of cantonese English, apparent native/nonnative status, and their effect on nonnative English speakers' perception. *Journal of Multilingual and Multicultural Development*, *30*(3), 253-269.

Huang, S. (1999). The emergence of a grammatical category definite article in spoken Chinese. *Journal of Pragmatics*, *31*, 77-94.

Hunt, K. W. (1970). Systactic maturity in schoolchildren and adults. *Monographs of the Society for Research in Child Development*, *35*(1, Serial No. 134).

Iverson, P. (2003). A perceptual interference account of acquisition difficulties for non-native phonemes. *Cognition*, *87*(1), B47-B57.

Jilka, M. (2000). *The contribution of intonation to the perception of foreign accent.* Unpublished doctoral dissertation, Universität Stuttgart.

Johnson, K. (1997). *Acoustic and auditory phonetics.* Malden, MA: Blackwell Publishing.

Kormos, J., & De´nes, M. (2004). Exploring measures and perceptions of fluency in the speech of second language learners. *System*, *32*(2), 145-164.

Kuhl, P. K., & Iverson, P. (1995). Linguistic experience and the ”perceptual magnet effect”. In W. Strange (Ed.), *Speech perception and linguistic experience: Theoretical and methodological issues* (p. 121-154). Timonium, MD: York Press.

Ladefoged, P. (1975). *A course in phonetics. (1 edition).* New York: Harcourt, Brace Jovanovich.

Ladefoged, P., & Maddieson, I. (1996). *The sounds of the world's languages.* Oxford: Blackwell Publishers.

Lee, T.-L., He, Y.-F., Huang, Y.-J., Tseng, S.-C., & Eklund, R. (2004). Prolongation in spontaneous Mandarin. In *Proceedings of the INTERSPEECH 2004 - ICSLP* (p. 526-529).

Lee, W.-S., & Zee, E. (2003). Standard Chinese (Beijing). *Journal of the International Phonetics Association*, *33*, 109-112.

Lenneberg, E. (1967). *Biological foundations of language.* New York: Wiley.

Lennon, P. (1990). Investigating fluency in EFL: A quantitative approach. *Language Learning*, *40*(3), 387-417.

Lev-Ari, S., & Keysar, B. (2010). Why don't we believe non-native speakers? the influence of accent on credibility. *Journal of Experimental Social Psychology*, *46*, 1093-1096.

Levelt, W. J. M. (1989). *Speaking: From intention to articulation.* Cambridge, MA: MIT Press.

Lin, C.-K., Tseng, S.-C., & Lee, L.-S. (2005). *Important and new features with analysis for disfluency interruption point (IP) detection in spontaneous Mandarin speech. Proceedings of Disfluency in Spontaneous Speech (DISS 05),*, 117-121. Aix-en-Provence, France.

Lin, Y.-H. (1989). *Autosegmental treatment of segmental processes in Chinese phonology.* Unpublished doctoral dissertation, University of Texas, Austin.

Lin, Y.-H. (2007). *The sounds of Chinese.* New York: Cambridge University Press.

Lindemann, S. (2000). *Non-native speaker "incompetence" as a construction of the native listener: Attitudes and their relationship to perception and comprehension of korean-accented English.* Unpublished doctoral dissertation, University of Michigan.

Lindemann, S. (2003). Koreans, Chinese, or Indians? Attitudes and ideologies about non-native English speakers in the United States. *Journal of Sociolinguistics*, *7*(3), 348-364.

Lindemann, S. (2005). Who speaks 'broken English'? US undergraduates' perceptions of non-native English. *International Journal of Applied Linguistics*, *15*(2), 187-212.

Long, M. H. (1990). Maturational constraints on language development. *Studies in Second language Acquisition*, *12*(3), 251-285.

MacKay, I. R. A., & Flege, J. E. (2004). Effects of the age of second language learning on the duration of first and second language sentences. *Applied Psycholinguistics*, *25*, 373-396.

Mohle, D. (1984). A comparison of the second language speech production of different native speakers. In H. W. Dechert, D. Mohle, & M. Paupach (Eds.), *Second language productions* (p. 50-68). Tubingen, German: Narr.

Montrul, S. A. (2006). On the bilingual competence of Spanish heritage speakers: Syntax, lexical-semantics and processing. *International Journal of Bilingualism*, *10*(1), 37-69.

Montrul, S. A. (2008). *Incomplete acquisition in bilingualism: Re-examining the age factor*. Amsterdam: John Benjamins.

Morrison, G. S. (2006). *L1 and L2 production and perception of English and Spanish vowels: A statistical modeling appraoch*. Unpublished doctoral dissertation, University of Alberta, Edmonton.

Munro, M. J. (1993). Productions of English vowels by native speakers of Arabic: Asouctic measurements and accentedness ratings. *Language and Speech*, *36*(1), 39-66.

Munro, M. J. (1998). The effects of noise on the intelligibility of foreign-accented speech. *Studies in Second Language Acquisition*, *20*(2), 139-154.

Munro, M. J., & Derwing, T. M. (1998). The effects of speaking rate on listener evaluations of native and foreign-accented speech. *Language Learning*, *48*(2), 159-182.

Munro, M. J., & Derwing, T. M. (2001). Modeling perceptions of the accentedness and comprehensibility of L2 speech. *Studies in Second Language Acquisition*, *23*, 451-468.

Munro, M. J., Derwing, T. M., & Morton, S. L. (2006). The mutual intelligibility of L2 speech. *Studies in Second Language Acquisition*, *28*(1), 111-131.

Papajohn, D. (1998). *Toward speaking excellent, the michigan guide to maximizing your performance on the TSE test and SPEAK test.* The University of Michigan Press.

Perkell, J. S., Guenther, F. H., Lane, H., Matthies, M. L., Perrier, P., Vick, J., et al. (2000). A theory of speech motor control and supporting data from speakers with normal hearing and with profound hearing loss. *Journal of Phonetics*, *28*, 233-272.

Perkell, J. S., Matthies, M. L., Svirsky, M. A., & Jordan, M. I. (1993). Trading relations between tongue-body raising and lip rounding in production of the vowel /u/: A pilot motor equivalence study. *Journal of Acoustical Society of America*, *93*(5), 2948-2961.

Peterson, G. E., & Barney, H. L. (1952). Control method used in a study of the vowels. *Journal of Acoustical Society of America*, *24*(2), 175-184.

Piske, T., MacKay, I. R. A., & Flege, J. E. (2001). Factors affecting degree of foreign accent in an L2: A review. *Journal of Phonetics*, *29*, 191-215.

Polinsky, M. (2006). Incomplete acquisition: American Russian. *Journal of Slavic Linguistics*, *14*, 191-262.

Ramsey, R. S. (1987). *The languages of Chinese.* Princeton, NJ: Princeton University Press.

Ramus, F., Nespor, M., & Mehler, J. (1999). Correlates of linguistic rhythm in the speech signal. *Cognition*, *73*, 265-292.

Read, J. (2000). *Assessing vocabulary.* Cambridge University Press.

Riggenbach, H. (1991). Toward an understanding of fluency: A microanalysis of non-native speaker conversation. *Discourse Process*, *14*, 423-441.

Shih, C. (1986). *The prosodic domain of tone sandhi in Mandarin Chinese.* Unpublished doctoral dissertation, University of California at San Diego.

Shih, C. (1995). Study of vowel variations for a Mandarin speech synthesizer. *Proceeding of EUROSPEECH-1995*, 1807-1811.

Shih, C. (2006). The language class as a community: A task design for speaking proficiency training. *Journal of Chinese Language Teachers Association*, *41*(2), 1-22.

Shih, C. (2008). Phonology and phonetics: An implementation model of tones. *Interfaces in Chinese Phonology*, 99-120.

Shih, C., & Ao, B. (1997). Duration study for the Bell laboratories Mandarin text-to-speech system. In J. van Santen, R. Sproat, J. Olive, & J. Hirschberg (Eds.), *Progress in speech synthesis* (p. 382-399). New York: Springer-Verlag.

Shih, C., & Lu, H.-Y. (2010). Prosody transfer: Tone production errors from second language learners. *Proceedings of the 5th International Conference on Speech Prosody 2010.*

Shih, C., & Wu, C.-H. (2011). Evaluating second language fluency. *Proceedings of New Tools and Methods for Very-Large-Scale Phonetic Research.*

Shriberg, E. (2001). To 'errrr' is human: Ecology and acoustics of speech disfluencies. *Journal of the International Phonetics Association*, *31*(1), 153-169.

Speer, S., Shih, C., & Slowiaczek, M. (1989). Prosodic structure in language understanding: Evidence from tone sandhi in Mandarin. *Language and Speech*, *32*(4), 337-354.

Stevens, K. N. (1972). *The quantal nature of speech: evidence from articulatory-acoustic data.* In E. E. Davis, Jr. Denes, & P.-B. Denes (Eds.), (p. 51-66). McGraw-Hill.

Stevens, K. N. (1989). On the quantal nature of speech. *Journal of Phonetics*, *17*, 3-45.

Stevens, K. N., & House, A. S. (1955). Development of a quantitative model of vowel articulation. *Journal of Acoustical Society of America*, *27*, 484-493.

Strange, W., Bohn, O.-S., Trent, S. A., & Nishi, K. (2004). Acoustic and perceptual similarity of north German and American English vowels. *The Journal of the Acoustical Society of America*, *115*(4), 1791.

Suter, R. (1976). Predictors of pronunciation accuracy in second language learning. *Language Learning*, *26*, 233-253.

Tepperman, J. (2009). *Hierarchical methods in automatic pronunciation evaluation.* Unpublished doctoral dissertation, University of Southern California.

Thomson, R. I. (2007). *Modeling L1/L2 interactions in the perception and production of English vowels by Mandarin L1 speakers: A training study.* Unpublished doctoral dissertation, University of Alberta, Edmonton.

Thomson, R. I., Nearey, T. M., & Derwing, T. M. (2009). A modified statistical pattern recognition approach to measuring the crosslinguistic similarity of Mandarin and English vowels. *Journal of Acoustical Society of America*, *126*(3), 1447-1460.

TOEFL iBT test: Independent speaking rubrics (scoring standards) (2008). http://www.ets.org/toefl/English/programs/scores/guides.

Torng, P.-C. (2000). *Supralaryngeal articulator movement in Mandarin Chinese tonal production..* Unpublished doctoral dissertation, University of Illinois, Urbana-Champaign.

Towell, R., & Dewaele, J.-M. (2005). The role of psycholinguistic factors in the development of fluency. In J.-M. Dewaele (Ed.), *Focus on french as a foreign language* (p. 210-239). Clevedon, UK: Multilinguall Matters.

Tremblay, A. (in press). Proficiency assessment standards in second language acquisition research: “clozing” the gap. *Studies in Second Language Acquisition.*

Tremblay, A., & Garrison, M. D. (2010). Cloze tests: A tool for proficiency assessment in research on L2 french. In M. T. Prior, Y. Watanabe, & S.-K. Lee (Eds.), *Selected proceedings of the second language research forum 2008* (p. 73-88). Somerville, MA: Cascadilla Press.

Trofimovich, P., & Baker, W. (2006). Learning second language suprasegmentals: Effect of L2 experience on prosody and fluency characteristics of L2 speech. *Studies in Second Language Acquisition*, *28*, 1-30.

Tseng, S.-C. (2003). Repairs and repetitions in spontaneous Mandarin. *Proceedings of Disfluency in Spontaneous Speech (DISS 03)*, 73-76.

Tseng, S.-C. (2006). Reparis in Mandarin conversation. *Journal of Chinese Linguistics*, *34*(1), 80-120.

Van Borsel, J., Sunaert, R., & Engelen, S. (2005). Speech disruption under delayed auditory feedback in multilingual speakers. *Journal of Fluency Disorders*, *30*, 201-217.

Vanderplank, R. (1993). Pacing and spacing as predictors of difficulty in speaking and understanding English. *English Language Teaching Journal*, *47*, 117-125.

Vorster, J. (1980). *Manual for the test of oral language production (T.O.L.P).* Pretoria, RSA: South African Human Sciences Research Council.

Wang, J. Z. (1993). *The geometry of segmental geatures in Beijing Mandarin.* Unpublished doctoral dissertation, University of Delaware.

Waves+ manual version 5.1. (1996). Entropic Research Laboratory, Inc.

Wilson, E. O., & Spaulding, T. J. (2010). Effects of noise and speech intelligibility on listener comprehension and processing time of korean-accented English. *Journal of Speech, Language, and Hearing Research*, *53*, 1543-1554.

Winer, B. (1971). *Statistical principles in experimental design (2nd ed.).* New York: McGraw-Hill.

Wode, H. (1983). Contrastive analysis and language learning. In H. Wode (Ed.), *Papers on language acquisition, language learning, and language teaching* (p. 202-212). Heidelberg, Germany: Groos.

Wu, C.-H. (2008a). A comparative study of L1 and L2 vowel quality. *Proceedings of the International Conference on Linguistic Evidence: Empirical, theoretical, and computational perspectives,*, 208-211.

Wu, C.-H. (2008b). Filled pauses in L2 Chinese: A comparison of native and non-native speakers. *Proceedings of the 20th North American Conference of Chinese Linguistics (NACCL-20)*, 218-228.

Wu, T.-C., & Lin, M.-t. (1989). *Shih yen yu yin hsueh kai yao*. Beijing: Kao teng chiao yu chu pan she.

Yoon, S.-Y. (2009). *Automated assessment of speech fluency for L2 English learners*. Unpublished doctoral dissertation, University of Illinois at Urbana-Champaign.

Youmans, G. (1990). Measuring lexical style and competence: The type-token vocabulary curve. *Style*, *24*, 584-599.

Yuan, J., & Liberman, M. (2008). Speaker identification on the scotus corpus. *Proceedings of Acoustics '08*.

Zhao, Y., & Jurafsky, D. (2005). A preliminary study of Mandarin filled pauses. *Proceedings of Disfluency in Spontaneous Speech (DISS 05)*, 179-182.